Assessing Geographic Relevance for Mobile Information Services

Dissertation

zur

Erlangung der naturwissenschaftlichen Doktorwürde
(Dr. sc. nat.)

vorgelegt der

Mathematisch-naturwissenschaftlichen Fakultät

der

Universität Zürich

von

Stefano De Sabbata

aus

Italien

Promotionskomitee

Prof. Dr. Sara Irina Fabrikant (Vorsitz)

Dr. Tumasch Reichenbacher (Leitung der Dissertation)

Prof. Dr. Ross Stuart Purves

Prof. Dr. Christoph Schlieder

Zürich, 2013

Printed version, Oxford, September 22^{th}, 2013.

Summary

In just twenty years, the World Wide Web and mobile networks have changed the way we communicate, the way we search for information we need, and the way we organize our activities. The opportunities afforded by these changes led to the development of literally millions of applications, including location-based services, which aim to provide information based on a user's position. However, the search methods used by such mobile information services have been criticized, because of the simplistic approach they employ.

Aim of this dissertation is to offer an in-depth study of the concept of relevance in mobile information services. The term "geographic relevance" has been proposed in literature to address such a concept, in order to emphasize the importance of space and time in assessing relevance in mobile information services. In this dissertation, I provide a detailed discussion of the concept of geographic relevance, offer a conceptual model of geographic relevance, and propose a list of novel criteria of relevance to be used in assessing the geographic relevance of geographic objects in mobile information services. Two empirical studies are presented, which confirm the validity of the proposed model and criteria. Based on these outcomes, I propose a computational method for the assessment of geographic relevance. The results show that the proposed assessment method is able to replicate human judgements of geographic relevance. Therefore, I argue that the proposed approach is an effective method for assessing geographic relevance in mobile information services.

The outcomes of the empirical studies presented in this dissertation confirm previous studies, showing how spatial proximity is not sufficient to consider a geographic object as relevant – assuming it is what a user is searching for. The temporal aspect plays a fundamental role, as temporally unavailable objects are to be considered as non-relevant. Moreover, in this dissertation, I suggest and prove that the geographic environment of a place or object has to be taken into account when assessing its geographic relevance. The presence of more objects of the same category in a neighborhood, or the presence of objects belonging to correlated categories, increases the relevance of the object. Geographic relevance is also dependent on the opportunities offered to a user in the neighborhood. Finally, as a user commonly seeks for places in order to perform an activity, I advocate the importance of taking into account the consequences of a user performing an activity in a given place, with respect to previous, concurrent, and subsequent activities, when assessing the geographic relevance of places or geographic objects.

Acknowledgements

I want to tank, first and foremost, Dr. Tumasch Reichenbacher, for giving me the great opportunity to pursue my Ph.D. at the Department of Geography, University of Zurich. Tumasch deserves most of the credit for the completion of my doctoral studies, for his unique contribution to my research and education, for all our thoughtful discussions, and for our friendship.

I would particularly like to thank Prof. Dr. Sara I. Fabrikant, for her precious advice on my research, and for guiding me through the conduction and analysis of the experiments presented in this dissertation. I am also grateful to have had the opportunity to collaborate with Prof. Dr. Ross S. Purves, and for all the valuable comments and reflections on this project. I would furthermore like to thank Prof. Dr. Christoph Schlieder for his helpful contribution and suggestions, and Prof. Dr. Christopher B. Jones for his scrupulous review of my dissertation.

I also wish to thank all my colleagues and friends at the Department of Geography, and in particular my fellow Ph.D. students at GIVA. I would like to acknowledge Paul Crease for his precious contribution to my research, and patient, kind help with all my writing. A special thank goes to Dr. Arzu Çöltekin for her support and friendship, for the projects we collaborated on, and the *"food for thought"*[1] .

Thanks to Prof. Dr. Stefano Mizzaro, for asking me whether I had ever considered doing a Ph.D., and for our thoughtful discussions on the concept of relevance. I would also like to mention Dr. Omar Alonso, for his valuable contribution to the crowdsourcing experiments presented in this dissertation.

Last, but by no means least, I acknowledge all my family. Giorgio and Nicolina, I owe them everything that I might ever accomplish. Massimo, for being better than me. Giulia, *"the only thing that really matters to me"*.

This thesis is dedicated to the memory of my grandmother, Regina "Gina" Gressani (20 November 1922 – 4 June 2013).

[1] `http://www.geo.uzh.ch/microsite/foodforthought/`

Contents

List of Figures

Chapter 1

Introduction

Computer-based information processing is becoming *"handy"*. In far less than a century, huge and complex number-crunching machines became folio-sized, multi-media devices, that even kids can easily interact with. In the '80s, computers entered people's homes with the introduction of the personal computer. These were still rather expensive, mostly professional devices, some of which were able to connect to the Internet through dial-up access services. By the first decade of the 21st century, the usage of mobile phones became popular, along with the broadband Internet access to the World Wide Web. Finally, due to the availability of mobile Internet connectivity, and the impressive increase of computational power of mobile devices, nowadays the Web is at everybody's fingertips. The scientific and technological shift has been wide-spread, changing how we do research, how we teach, how we travel, how we communicate, and how we make decisions.

The focus of this dissertation is strongly related to how the advances mentioned above can change our decision making processes, in particular when we search for places on the go. A prototypical scenario is a person using her mobile phone to seek for a place nearby (e.g., a restaurant, a hotel, a supermarket, or a first aid centre) in order to serve a need or accomplish an activity. This dissertation investigates the new possibilities available to facilitate such information seeking tasks, new research questions, possible issues related to naïve approaches, viable solutions, and foreseeable advancements.

1.1 Motivation

The World Wide Web is the source many people go to when searching for information (Fallows, 2004) when they have nobody to ask. Especially when on the go, the mobile phone is the device that many people have almost always at hand. It seems quite straightforward that if someone has an information need while on the go, one option is to pick up the mobile phone, open the browser or another application, and retrieve information through the Internet. This has been found to be a valuable choice for information seeking for over 80% of smartphone users in Europe, the U.S., and Japan[1]. As a consequence, in the last decade both the Web and the mobile application market have seen a proliferation of applications providing location-based and geographic-related information.

Showing static maps in the form of images has been common practice since the early stages of the Web, especially for business websites and news agencies (Jones and Purves, 2008b). The first service offering interactive maps on the Web was the Xerox PARC Map Viewer (Putz, 1994; Towers and Gittings, 1995) developed at Xerox Corporation's Palo Alto Research Center[2]. Starting in 2001, ViaMichelin[3] was the first to offer both interactive maps and routing services on the Web, and also as application for personal digital assistants – the predecessors of smartphones. However, such services did not reach a wide audience until Google entered the Web-mapping field with the Google Maps[4] service in 2005, and Apple shook the mobile phone market with its iPhone[5] in 2007. Nowadays, all new mobile phones are equipped with satellite-based positioning systems – e.g., the Global Positioning System (GPS)[6] – and have map applications.

In the same years, three other developments were changing the Web. The first one was the open source[7] movement. Open, collaborative, and free generation and access to information are the bases of this movement, which have given us important projects such as the user-generated encyclopaedia Wikipedia[8] and the operating system Linux[9]. Similar initiatives have also been promoted concerning geographic data (e.g., OpenStreetMap[10]) and are commonly referred to as volunteered geographic information (Goodchild, 2008). In the scope of such projects, each user can submit new content and modify existing information, and each user can access all the produced content.

The second (and perhaps less known) one is the open data movement, originating with a campaign promoted by the Guardian Technology[11] in the U.K. in 2006. Its aim is to make the data produced by government agencies freely available to the public. The idea is that citizens pay for the collection and analysis of the data with their taxes, and they should be able to freely access the data. Despite the fact that this was already common practice in the U.S. (Jones and Purves, 2008b), such was not the case in Europe. In December 2011, the European Commission launched the Open Data Strategy for Europe as part of its Digital Agenda for Europe[12], adopting as a general rule that all documents and data should be offered for free and can be re-used for any purpose.

[1]Source: Google/MMA, Global Perspectives: The Smartphone User & Mobile Marketer, June 2011. `http://www.thinkwithgoogle.com/insights/library/studies/`, last accessed September 2012.
[2]`http://www.parc.com/`, last accessed September 2012.
[3]`http://www.viamichelin.com/`, last accessed April 2012.
[4]`http://maps.google.com/`, last accessed April 2012.
[5]`http://www.apple.com/iphone/`, last accessed May 2012.
[6]`http://www.gps.gov/`, last accessed May 2012.
[7]`http://en.wikipedia.org/wiki/Open_source`, last accessed April 2012.
[8]`http://www.wikipedia.org/`, last accessed April 2012.
[9]`http://en.wikipedia.org/wiki/Linux`, last accessed April 2012.
[10]`http://www.openstreetmap.org/`, last accessed April 2012.
[11]`http://www.guardian.co.uk/theguardian/technologyguardian/technology`, last acc. April 2012.
[12]`https://ec.europa.eu/digital-agenda/`, last accessed November 2012.

The third one is probably one of the most controversial, widespread, well known, and little understood subject of the last years: social networks. A huge flow of continuously produced data, increasingly attached with location-related information. This abundance of previously scarce social data generated the field of Big Data (e.g., boyd and Crawford, 2011), which aims to understand the complex relations hidden within them.

Social and technological changes have made obsolete the idea of a service provider with a stable dataset of points of interest to be displayed on a map, printed out for a given usage purpose. Nowadays, maps can be digital, interactive, mobile, and constantly updated with data from a number of diverse sources: from a proprietary dataset to OpenStreetMap, from your friends to government agencies. Nevertheless, this vast amount of information is not just a resource, but can also be an issue. In the described scenario, it would not be feasible to show all the available geographic information on a map, especially on a mobile phone. Information overload, inaccurate relevance assessment, visual cluttering, and ambiguous representation can slow down or hinder users' decision making processes, or mislead users (Fischer, 2001; Reichenbacher, 2004, 2009; Crease and Reichenbacher, 2011).

The challenge is then, how can an information system extricate itself in this jungle of geographic information, and retrieve the most relevant information for the user in a given context? Future services and applications will have to understand the context in which a user is using a map or submitting an information request. This will allow them to filter out less relevant information and represent geographic entities according to their relevance.

In order to handle this complex problem, a new concept of relevance (Zipf, 2003; Reichenbacher, 2005a; Raper, 2007; Reichenbacher et al., 2009; Reichenbacher and De Sabbata, 2011) has emerged in geographic information science (GIScience), which has been named "Geographic Relevance" (GR) by Raper (2007). GR refers to the relevance of a geographic entity (i.e., a physical entity or feature in the real world), given a specific context of interacting with its representation, such as a point of interest on a digital map. It is important to underline that even if this concept is strongly related to the concept of relevance used in information retrieval (IR), GR is not a property of a document, but refers to the relevance of a real world entity. Geographically-referenced documents and documents reporting geographic information can be used as a source of information in order to judge the GR of an entity, but they are not the objective of the relevance assessment. A similar concept was also suggested within the field of IR by Coppola

Definition of Geographic Relevance

"We define GR as a quality of an entity in geographic space or its representation in an information system, i.e. an object, document, or image. This quality is expressed as the relation between an entity or its representation and the actual context of using the representation." (Reichenbacher and De Sabbata, 2011, p. 68)

et al. (2004). Their conceptualisation of relevance was based on the notion of situational relevance (Wilson, 1973; Saracevic, 2007) and referred to the relevance of objects in the physical world with respect to the user context.

1.2 Thesis objectives

The concept of GR can be a fundamental tool for future mobile geographic information services to use to help avoid information overload and related issues. For this purpose, it is then necessary to define an operational assessment method for numerically calculating GR. The outcome of such a procedure is a numerical value estimating the GR for a given geographic entity in a given context.

The aim of this dissertation is to first formalise the concept of GR and thereafter define a method for its quantitative assessment. I applied a top-down approach, investigating the diverse facets of GR (e.g., the thematic and spatial ones), analysing the diverse entities involved in it (i.e., the user and the geographic entities) and their components. On that basis, I proposed and experimentally validated an assessment method, which numerically estimates the strength of each of those facets and combines the resulting values into a single GR score. The main research question pursued is:

- *Which information and criteria are needed, and how do they have to be combined in order to assess a set of numerical values that estimate the GR of a given geographic entity with respect to a given context of use?*

Modelling

GR has been defined as a relationship between the context of a mobile user and a geographic entity in her environment. In order to be reasonably tractable by a computer algorithm, this complex relationship has to be broken down into simpler relationships. These relationships will involve pieces of information describing the user's context on the one side, and pieces of information describing the geographic entity on the other side. These include the user's interest and current activity, but also her position, or time schedule. The category of geographic entities and offered services at places also have to be considered, along with their location and time validity. The user's knowledge of the environment may also play an important role, as well as the environment surrounding the geographic entity and its relationships to other entities nearby. The first research question is then:

- **RQ1**: *Which information is needed to model the user context, the geographic entities, and the surrounding environment in order to assess GR?*

Once all key information has been identified, the relationship between the user context and the geographic entities in its components can be inspected. Analogue to the relationship between a user query and the content of a webpage, which is employed as topicality criterion in web search engines, each single relationship between a piece of information describing the user's context and a piece of information describing a geographic entity can be seen as a criterion of GR. The second research question is then:

- **RQ2**: *Which criteria can be used to assess the relevance of geographic entities?*

Assessment

If GR is distinct from the concept of relevance commonly used in IR, then the answer to the second research question (RQ2) will include new criteria, which have to be taken into account to assess GR, but which are not taken into account by current methods in IR. Thus, for each one of these criteria, a formal method of computation has to be established, which will calculate the strength of the relationship between the user's context and the geographic entity for that particular component. The third research question is then:

- **RQ3**: *How can we use information about user context, geographic entities, and the surrounding environment to compute a numerical value for each criterion of GR?*

Finally, once a method has been established to calculate a numerical value for each identified criterion, it will be necessary to aggregate these values in a single relevance value to avoid cognitive overload issues for users. In fact, it is most likely that an application would need such an aggregated value to communicate the GR of a geographic entity to the user. This would be the case if a mobile information system presents GR as a rank, and also if the information is presented on a map – using a visual variable (Bertin, 1983) to represent the relevance of the geographic entity. More sophisticated representations showing the user more than one facet of GR are also imaginable, but they would be far more complicated to understand and thus less likely to be implemented in a mobile information system. The final objective of this dissertation is to suggest a possible solution to the problem of how to aggregate the values related to the single criteria. The fourth research question is then:

- **RQ4**: *How can we combine the values representing single criteria of GR in an adequate way?*

1.3 Thesis structure

The remainder of the thesis is organised as follows (see Figure 1.1). Chapter 2 reviews the role of relevance in computer science and geography, and the fields and concepts that will be discussed in this dissertation to deepen the study of GR. The latter is examined in detail in Chapter 3, which presents the conceptual framework of my research. Chapter 4 reports on the two experiments presented in (De Sabbata and Reichenbacher, 2012), which investigate the criteria of geographic relevance and their importance. The outcome of these two experiments is thereafter rendered into the main GR assessment method *ScoreGR*, which is presented in Chapter 5, along with an alternative GR assessment method referred to as *GRBM25*. Chapter 6 reports the details related to the prototype implementation of both GR assessment methods. Chapter 7 presents the third and last experiment reported in this dissertation, which was performed in order to evaluate the GR assessment methods as proposed and implemented in the two previous chapters. Chapter 8 discusses the effects of this evaluation on the concept, criteria, and assessment methods of GR. Chapter 9 concludes this dissertation, summarising the overall findings and contribution of my research project, and identifying potential future research directions in this field.

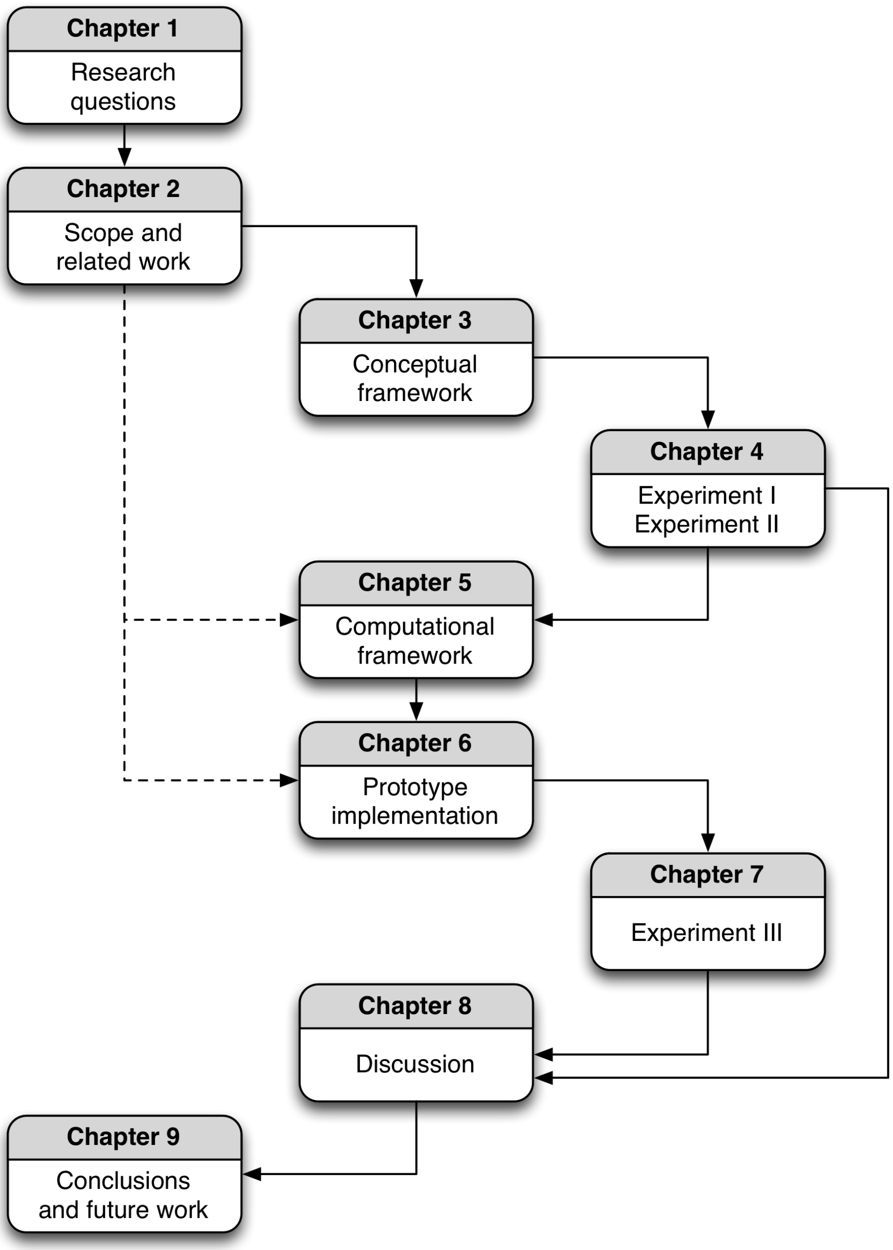


Figure 1.1: Thesis workflow.

Chapter 2

Scope and related work

The background of this work is both in geographic information science (GIScience) and information retrieval (IR). GIScience comprehends *"the development and use of theories, methods, technology, and data for understanding geographic processes, relationships, and patterns. The transformation of geographic data into useful information is central to geographic information science"* (Mark, 2003, p. 1; Goodchild, 2012, p. 6). IR is *"finding material (usually documents) of an unstructured nature (usually text) that satisfies an information need from within large collections (usually stored on computers"* (Manning et al., 2008, p. 1).

Both IR and GIScience experienced a great spread through the Web, but they were still some years away from the mobile and social networks revolutions, when Zipf (2003) first suggested taking into account the concept of relevance when displaying geographic entities on a computer map. Shortly afterwards, Reichenbacher (2004, 2005a) showed how the integration of such a relevance concept in the design of mobile maps can enhance their usability. Finally, Raper (2007) suggested the term '*geographic relevance*' (GR) to refer to the relation between the user's need for geographic information, associated to an activity performed or planned in a geographic context, and the spatial, temporal, and utilitarian (functional) expression of entities in the geographic space.

As a Web document can be relevant to an information need (Cooper, 1971; Manning et al., 2008) of a user of an IR system, geographic entities (such as shops, hotels, and hospitals) can be relevant to an activity a user is performing or planning, at a given time, in a given location and surrounding environment. In fact, in presenting a review of location-based applications for current mobile operating systems (e.g., Android[1], iOS[2], and Windows Phone[3]), Hauthal and Burghardt (2012) show how most of these applications have a spatial search function.

Section 2.1 starts with the concept of relevance and its evolution within the field of IR. Section 2.2 is devoted to the concept of context, to its main components of location, mobility and human activity, and to how this concept is fundamental in understanding

[1] `http://www.android.com`, last accessed March 2012.
[2] `http://www.apple.com/ios`, last accessed March 2012.
[3] `http://www.microsoft.com/windowsphone`, last accessed March 2012.

relevance. Section 2.3 covers the core subject of this dissertation with a review of research dealing with the relationship between geography and relevance, in both IR and GIScience literature. Section 2.4 and Section 2.5 give an account of geographic data mining and information analysis methods and system evaluation procedures, which will be referred to later on in this dissertation.

2.1 Relevance

As reported by Reichenbacher (2007), the term relevance has its origin in the Latin verb meaning to rise or to leave. The term was first used in law to describe the legal sufficiency or adequacy of a claim (Oxford English Dictionary). Figuratively this is understandable by the use of a pair of scales to weigh the arguments. In a general sense, relevance means a connection with the subject or point at issue, a relation to the matter in hand (Oxford English Dictionary).

Greisdorf (2000) defined relevance as '*the criterion used to quantify the phenomenon involved when individuals (users) judge the relationship, utility, importance, degree of match, fit, proximity, appropriateness, closeness, pertinence, value or bearing of documents or document representations to an information requirement, need, question statement, description of research, treatment, etc*'. This definition clearly highlights how relevance is still a fuzzy and ill-defined concept, which also entails different meanings in various disciplines, such as philosophy, psychology, linguistics, computer science, and information science.

The term relevance has been used since the 1950s in information science in general and in IR in particular to refer to the relation of pertinence of a document to a given information need, or its appropriateness to the matter a user is interested in (Manning et al., 2008). This notion is crucial in the definition of situational relevance proposed by Wilson (1973), which is discussed in further details in Section 3.1, as it is at the core of current conceptualisations of relevance, including the concept of geographic relevance (Reichenbacher and De Sabbata, 2011) discussed in this thesis.

In the 1960s, the first IR systems applied a binary relevance, based on term matching between query and documents. The documents containing one or all the terms specified in the user's query were considered as relevant, the rest were considered as irrelevant. Despite the advance in many aspects of the retrieval process, binary relevance was the most widely used concept in IR, until recent claims for a multidimensional and more fuzzy concept (Saracevic, 1996; Cosijn and Ingwersen, 2000). Current document-based IR systems still base their ranking on the correspondence or agreement between the terms specified in a query and the terms contained in a document. This measure is combined with a measure of the popularity of the same document, the authoritativeness of the source, and many other criteria. For further details, Mizzaro (1997b) presents a comprehensive history of the development of the concept of relevance in IR and its implementation in academic systems. Few details are actually publicly available about

Definition of Relevance

"We suggest that relevance be defined as a relation between an individual, at the time he senses a need for information, and a document. We shall say the document is relevant to that person if he feels the need that brought him to examine the document are satisfied, at least in part." (Bookstein, 1979, p. 268)

"An information need is the topic about which the user desires to know more. (...) A document is relevant if it is one that the user perceives as containing information of value with respect to their information need." (Manning et al., 2008, p. 5)

the ranking systems used in current Web search engines such as Bing[4], Google[5], and Yahoo![6].

The binary term matching between query and documents can be considered as a basic understanding of the criterion topicality, which is defined as the extent to which a piece of information (usually a digital document stored in an archive) concerns the topic the user is interested in. Even the very first Web search engines were based on the same fundamental criterion, and the list of retrieved relevant Web pages was presented ordered alphabetically by title. This simple ranking method may work fine with a small set of documents. As soon as digital documents collections started growing, that was not the case anymore. Such a system would be almost useless if applied to the Web, which is currently estimated to contain as many as 50 billions pages[7] – considering that the estimation mentioned by Page et al. (1999) when presenting the Google's Page Rank algorithm was referring to a number of 150 million Web pages, only 14 years ago.

The first system to interpret a document as a vector in a term-defined space was probably the SMART IR system at Cornell (Salton, 1971). Such models conceptualise a space with as many dimensions as the number of distinct terms contained in the documents of a collection. In this space, an empty document would be a point in the origin. Any other document can be represented by a vector, having weight zero in each dimension referring to a term it does not contain, and a not-null weight for each dimension referring to a term that it contains. In the same space, with the same procedure, it is possible to represent a query. The relevance score of a document for a given query will then be derived by the dot product of the two vectors.

The main issue then is how to weight each document in each dimension. The main innovation in this field has been proposed by Spärck Jones (1972). She defined the inverse document frequency (idf) as the inverse measure of the number of documents in a collection that contain a given term. Given a query containing a set of terms, the idf is commonly used together with the term frequency (tf) – that is, a measure of occurrences

[4]`http://www.bing.com/`, last accessed March 2012.
[5]`http://www.google.com/`, last accessed March 2012.
[6]`http://search.yahoo.com/`, last accessed March 2012.
[7]`http://www.worldwidewebsize.com/`, last accessed March 2012.

of a given term in a given document – to estimate the weight of a document in a given dimension (i.e., for a given term). The less frequent the term is in the collection, the higher the score. The more frequent the term is in the document, the higher the score. Many formulas have been suggested which use tf and idf to compute weights. The most famous and frequently used tf-idf weighting schema is the Okapi BM25 (Spärck Jones et al., 2000) – named after the first system to implement it. The Okapi BM25 originated from the probabilistic IR approach (Maron and Kuhns, 1960; Robertson and Jones, 1976) and the binary independence model (van Rijsbergen, 1979).

Nevertheless, Bookstein (1979, p. 270) pointed out that '*relevance is not necessarily the same as topicality; a document on a different topic might, for one reason or another, satisfy the user's information need. Conversly, a document may not be judged satisfactory, if, for example, the patron is already familiar with its contents, or is interested in an aspect of the topic other than that treated in the document*'. Great effort has since been devoted to study the criteria people actually apply in information seeking, and a large number of additional criteria have been proposed.

The pioneer studies on the criteria of relevance (other than topicality) were conducted in the early 1990s. Schamber (1991) was the first to aim for criteria derived from the observation of user behaviour in information seeking tasks. She studied the criteria applied by users of a weather information system to evaluate the usefulness of presented information. This was the first study to focus on the judgement of the relevance of information retrieved by an information system. The elicited criteria were subdivided into ten categories: accuracy; currency; specificity, detail, and concreteness; the geographic proximity of the weather-related event to the user's location; the reliability of the source; the accessibility to information (including the ease of obtaining information, and costs); the verifiability of information through other sources; clarity of presentation; dynamism of the presentation; and presentation as source of entertainment and affective response for the user.

Though having similar purposes, a previous study by (Halpern and Nilan, 1988; Nilan et al., 1988) was more focused on the judgement of the relevance of information gathered from other persons, in particular in the case of health-related information needs. In this study, the participants mentioned criteria such as information coverage, expertise, experience, confirmation, uncertainty, human relationships criteria – such as relationship of the source to the user, friendliness, trust, respect, power, social pressure, confidentiality – financial and time criteria, and other criteria related to the healthcare method under consideration.

Following the same line of research, Barry (1994) studied the criteria taken into account by 18 students and faculty members in different fields of research. The scenarios were designed as an online search for information for preparing undergraduate level class assignments, graduate level class assignments, master's theses, doctoral dissertations, and professional presentations and publications. Barry's study reports 23 categories of criteria, most of which trace, specify, or generalise those mentioned above. Among the criteria identified exclusively by Barry are effectiveness, consensus within the field, time constraints (i.e., whether a document could be obtained in time for whatever deadline or not), ability to understand, and novelty. At the same time, Barry's study did not mention the criteria geographic proximity, dynamism and presentation quality, which were probably tightly related to the specificity of the weather information system taken into account by Schamber.

A detailed comparison of Schamber's and Barry's studies can be found in (Barry and Schamber, 1998), including a list of 10 joined criteria of relevance, which has been the base of all further research in this field. The list is reported in Table 2.1. In this jointly published study, the authors conclude that the major similarities in the two previous studies advocate for '*the existence of a finite range of criteria that are applied across types of users, information problem situations, and information sources*', whereas the few differences '*appear to be due to the differences in situational contexts and research task requirements*' (Barry and Schamber, 1998, p. 234).

First suggested by Barry (1994), the criterion novelty has been put forward again by Xu and Chen (2006) as part of a five-factor model of relevance including also topicality, reliability, understandability (i.e., clarity), and scope (i.e., specificity). The results of a study from Luyt et al. (2008) suggest that users first perceive the novelty of a document and then consider whether it is on topic or not. Nevertheless, the novelty of a document does not improve the relevance if it is not on topic. That means that novelty is a non-compensatory criterion – i.e., novelty can not compensate for a lack of topicality of a document, which would be then judged as not-relevant.

Similar to novelty, the criterion familiarity suggested by Savolainen and Kari (2006) refers instead to the extent to which the user is familiar with the source of information. In a broader sense, a parallel can be also drawn to the criteria related to human relationships as suggested by Nilan et al. (1988). In this case, the familiarity refers mostly to an institution or a company. In the same study, Savolainen and Kari (2006) put forward two other criteria. One is related to the variety of information provided by the source about a given topic, and another is related to the extent to which access to information is dependent on personal curiosity of the user.

In (da Costa Pereira et al., 2009), the authors focus their attention towards the user's preferences. On one hand, the criterion appropriateness is suggested, which refers to the extent to which the affordance (Jordan et al., 1998) of the entity is focused on the user's needs. On the other hand, the criterion coverage is defined as the extent to which the user's needs are satisfied by the affordance of the entity.

In Web search engines, one further criterion is considered which takes advantage of the hyperlinked structure of the Web (Marchiori, 1997) – this was usually not the case for document collection until the introduction of the Web (Berners-Lee, 1992) – and it is at the core of Web-based search strategies such as Google's Page Rank algorithm (Page et al., 1999). This criterion can be referred to as popularity. The underlying idea is that the more a page is linked by important pages, the higher its importance.

In most of the studies mentioned above, the implied scenario assumes that the user would interact with the IR system on a desktop computer by entering a textual query. As already mentioned in the introduction, nowadays this is not the only common scenario. Some of the criteria listed by Barry and Schamber (1998) already suggest that a system needs a better understanding of the context in which the query has been formulated in order to properly understand the user's information needs. This is even more important in the mobile use case. The following section therefore investigates concept of context.

Table 2.1: The ten criteria of relevance published by Barry and Schamber (1998).

Criterion	Description
Depth/Scope/Specificity	The extent to which information is in-depth or focused; is specific to the user's needs; has sufficient detail or depth; provides a summary, interpretation, or explanation; provides a sufficient variety or volume.
Accuracy/Validity	The extent to which information is accurate, correct or valid.
Clarity	The extent to which information is presented in a clear and well-organized manner.
Currency	The extent to which information is current, recent, timely, up-to-date.
Tangibility	The extent to which information relates to real, tangible issues; definite, proven information is provided; hard data or actual numbers are provided.
Quality of Sources	The extent to which general standards of quality or specific qualities can be assumed based on the source providing the information; source is reputable, trusted, expert.
Accessibility	The extent to which some effort is required to obtain information; some cost is required to obtain information.
Availability of Information/Sources of Information	The extent to which information or sources of information are available.
Verification	The extent to which information is consistent with or supported by other information within the field; the extent to which the user agrees with information presented or the information presented supports the user's point of view.
Affectiveness	The extent to which the user exhibits an affective or emotional response to information or sources of information; information or sources of information provide the user with pleasure, enjoyment or entertainment.

2.2 Context

The diffusion of mobile connection to the Internet through increasingly computationally powerful mobile devices has produced new ways to access and produce information. At the same time most mobile devices are equipped with rather precise and inexpensive chips for receiving satellite positioning systems signal. The result has been a fast growing amount of information delivered to and sent from various locations, mostly concerned with the same locations or activities that people perform in those locations.

At the same time, this presents an opportunity for applications to use that location information and communication stream to deliver innovative services, based on the user's location. This also presents novel issues related to the amount of information delivered, which can produce information overload, and consequently make those applications not usable. This is having a great impact in all fields of information science. In order to reduce the amount of delivered information and lower the risk of information overload, the context of information need can be taken into account. If the system is able to perceive the context in which a request is made, it can better understand the user's need, focus the scope of the information search, and deliver a smaller amount of more focused information.

The concept of context has been widely studied in computer science (for a review see Hong et al., 2009), but it still lacks a universally-accepted definition (as the concept of relevance in IR). Many applications use a relatively simplistic and *ad hoc* definition of context to offer specific services of adaptive personalised information delivery (e.g., Kapitsaki et al., 2009; Emmanouilidis et al., 2012). The aim of such applications is to adapt *"according to the location of use, the collection of nearby people, hosts, and accessible devices, as well as to changes to such things over time, [to] examine the computing environment and react to changes to the environment"* (Schilit et al., 1994, p. 85). In the IR domain, context-aware retrieval systems have been designed, where context was first understood as the set of topics collected from previous queries (Jones and Brown, 2004; Shen et al., 2005). More recently, the idea of a context-aware browser (Coppola et al., 2010; Mizzaro and Vassena, 2011) has be put forward as a new paradigm for context-aware mobile information access. In this case, commonly used low-level information coming from electronic sensors (e.g., GPS for the position) is joined with high-level information introduced by the user in the form of text-based tags. Both types of information are used to infer a context to be used in the IR process, which can also be proactive.

Definition of Context in ubiquitous computing

According to a widely cited definition, *"context is any information that can be used to characterise the situation of an entity. An entity is a person, place, or object that is considered relevant to the interaction between a user and an application, including the user and applications themselves."* (Dey, 2001, p. 5)

2.2.1 Location

Among all the factors which have been considered as possible components of context, location is one of those most frequently taken into account, at least in mobile applications. This is the case for location-based services (LBS), which are mobile applications that aim to offer simple spatial information processing capabilities on the base of the user's location (Shiode et al., 2002).

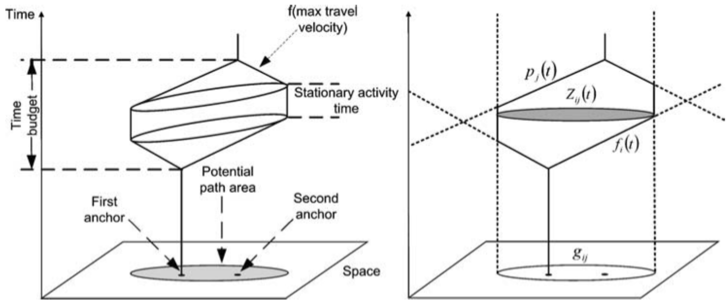
LBS implement the core concept of context-aware applications, and they are also referred to as location-aware applications. These services commonly employ a straightforward spatial filter to the information at hand, pruning all the objects which are further away than a fixed threshold. This can be seen as the spatial equivalent of the binary concept of relevance adopted by the first IR systems. LBS have the credit of being the first to be deployed to the broad public, due to their usefulness and easiness. Nevertheless, it is clear that mobile applications cannot limit their understanding of context just to the user's location (Schmidt et al., 1999; Reichenbacher, 2005b; Jiang and Yao, 2006; Raper et al., 2007b).

(Raubal et al., 2004) propose a conceptual framework for a user centred theory of LBS, that aims to take into account not only spatial, but also temporal, social, and cognitive aspects of users' preferences. An example of this approach is the adaptive mobile touristic application presented by Bereuter et al. (2009), which exploits user's preferences extracted from online social networking profiles in suggesting interesting places to visit. More recently, Abdalla and Frank (2012) proposed a formalism able to combine the routing functionalities of a LBS and the task-planning capabilities of a personal information management tool, in order to produce instructions that can lead the user to the fulfilment of a task by traveling between different locations. Investigating the requirements for the prototype service Google Hotel Finder, Riegelsberger et al. (2012) highlight the complexity of the trade-offs between quality, price, and location during the hotel search decision-making process. These are first steps to bring the field of LBS into the wider discussion concerning the concepts of context and relevance, which is the perspective adopted in mobile cartography (Reichenbacher, 2004) – as further discussed in Section 2.3.2.

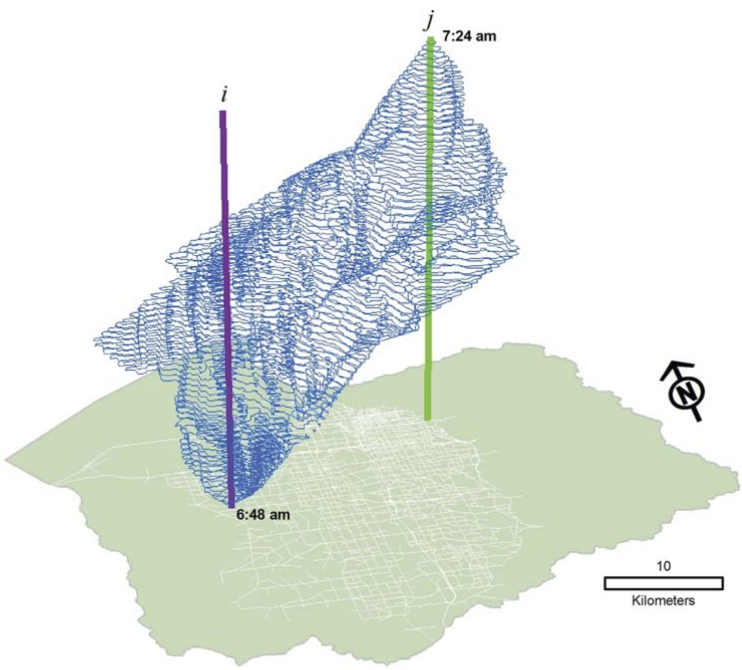
2.2.2 Mobility

Time geography (Hägerstrand, 1970; Pred, 1977) is a framework which has been developed within geography to study how human activities take place in space and time. Being a constraints-oriented approach, it defines the necessary conditions for a person to perform an activity based on spatial and temporal constraints. In fact, users have to 'trade' time for space in order to generate mobility (Miller and Bridwell, 2009).

The main tools developed within the time geography measurement theory for the analysis of these constraints are the space-time path and the space-time prism (Hägerstrand, 1970; Miller, 2005a,b). The space-time path describe the movement and the activities of an individual in space with respect to time. This is usually represented as a line in a three dimensional space (i.e., two spatial dimensions and one time dimension), connecting all the points where the individual's position has been measured with a related timestamp. The space-time prism is a definition of the potential mobility of the individual, given her current position in space and time, and possible other constraints. This is commonly represented as a prism in a three dimensional space (i.e., two spatial dimensions and one time dimension), containing all the possible locations that could become part of the space-time path in the case of a period of potential or unobserved travel (see Figure 2.1a).



(a)



(b)

Figure 2.1: A space–time prism in a Euclidean space (a) and in a field-based space representations (b) (Miller and Bridwell, 2009).

Space-time paths and space-time prisms are common representations of individual movements. They offer an efficient and simple computational framework, where the space is conceptualised as a two-dimensional Euclidean space, with no further constraints. This is currently the most used model in LBS, because of its low computational complexity. This can be crucial when the information has to be delivered in a matter of seconds, and a large number of options have to be taken into account as possible intermediate destinations. Nevertheless, network and field-based tools for time geography have been developed by Miller and Bridwell (2009) (see Figure 2.1b). Though they require a longer time for computation, these tools can give a better insight into user's mobility, especially in an urban environment.

More recently, Kuijpers et al. (2011) proposed the kinetic space-time prisms. The aim of this work is to push the realism of the model further by avoiding non-realistic assumptions, such as the individual's ability to instantaneously change direction and speed. Their work suggests that taking into account upper bounds of acceleration can have a significant effect on the outcome of the simulation and affect the geometry of the space-time prism. This would then change any calculation based on the space-time prism and affect the information presented to the user, and eventually her decision making process.

Although time geography was developed in the 1970s, only the recent proliferation of GPS-equipped mobile devices has enabled it to be applied to everyday activities. The concepts and formal models developed in the field of time geography can be used to define the mobility of the user, with respect to the time needed by the user to perform an activity she is performing or planning. That is, it can be calculated in which places a user will be able to perform a given activity, taking into account her location, time schedule, and mode of transportation. Thus, the representation of space offered by the field-based perspective of time geography can be a key factor in future intelligent transportation systems and LBS. In fact, this could be combined with a high-resolution data collection and analysis to achieve a more realistic model of users' potential mobility – i.e., a highly-detailed description of the locations the user would be able to visit within given spatial and temporal constraints. These are fundamental facets that should be at the core of any LBS, but which are still not implemented in most of the currently available applications. The integration of such methods into mobile document-based IR systems (Mountain, 2005; Mountain and Macfarlane, 2007) is later discussed in Section 2.3.1.

2.2.3 Activity

The majority of modern mobile phones is equipped with sensors sufficient to determine the user's position in space and time and with computational power to estimate user's potential mobility, at least at a basic level. This is fundamental information, but it is still not enough to fully understand the user's context. A key determinant of the context is the user's activity, whose central role as a context factor has been pointed out in (Crowley et al., 2002). Human activity is a complex construct, which is not easily derived from electronic sensors, and usually requires some kind of human-generated input to be used in context-aware applications.

Recent human-computer interaction studies have suggested the use of activity theory (Kaptelinin and Nardi, 1997, 2006) as a framework for modelling human activities in a context model (Nardi, 1996; Greenberg, 2001). Activity theory was developed in early 1920s by several Russian psychologists, Vygotsky, Rubinshtein, and Leont'ev among the others. This framework was developed to conceptualise human activities, through a formal model involving the main elements of activity – namely subject, tools, community, division of labour, rules, object, and outcome.

Kaenampornpan and O'Neill (2004) first suggested to incorporate such a model of activity in context models for context-aware and ubiquitous computing. A description of the user's context through the model defined in activity theory can help the modeller to identify all key elements of context that can influence a context-aware application. Shortly after, Reichenbacher (2005b) and Dransch (2005) recognised the applicability of activity theory to mobile services dealing with geographic information (see Section 2.3.2). Since mobile services differ from desktop-based geographic information services, they should take the user's activity into account and adapt to it.

In the field of LBS, Huang and Gartner (2009) suggested a mapping of the elements commonly used to describe the key elements of an activity according to activity theory to a set of context categories, which follow the traces of the five types of contexts defined in the taxonomy proposed by Kofod-Petersen and Cassens (2006). The proposed categorisation is '*inherited from the context-aware tradition and adopted to make use of the general concepts we find in activity theory*' (Kofod-Petersen and Cassens, 2006, p. 12) in the study of pedestrian wayfinding activities. The environmental, personal, and social context are respectively enfolding the user's surroundings, state of mind, and social aspects – such as '*information about the different roles a user can assume*', although nowadays it would be more closely related to social-network driven information. The task context incorporates an activity model derived from the activity theory. Finally, a spatio-temporal context is dedicated to location- and time-related information. The fundamental idea is to derive a general context from basic sensor input, and adapt LBS to the identified context, in a bottom-up fashion.

Using a top-down approach, Hirtle et al. (2011) showed how the user activity plays a central role in how people give route directions. The user's activity can be directly related to the granularity used for navigation, the inclusion of relevant details, and the exclusion of irrelevant details. Route directions given for emergency calls, are probably quite different from those given for touristic tours.

From the content of this section it becomes clear how relevance is dependent on context and how a substantial part of the context is geography, as well as events and phenomena taking place in geographic space. The relationship between geography and relevance is thus the subject of the next section.

Definition of Geographic information retrieval

"Geographic Information Retrieval, as we define it here, is an applied research area that combines aspects of DBMS research, User Interface research, GIS research, and Information Retrieval research, and is concerned with indexing, searching, retrieving, and browsing of georeferenced information sources, and the design of systems to accomplish these tasks effectively and efficiently." (Larson, 1996, p. 85)

2.3 Geography and relevance

Topicality measures used in current document-based IR systems integrate some degree of semantics. Synonym-based query enhancement and co-occurrence-based term disambiguation are common methods that IR systems use to improve their understanding of the user's need that lies behind a textual query Voorhees (1994); Xu and Croft (1996). Nevertheless, these methods may not always be enough to understand the explicit or implicit spatial references contained in a document. This can lead to poor results when dealing with queries that involve geographic information.

Geographic information retrieval (GIR) is the field that studies how to mine unstructured collections of documents for spatial references to be later used in the retrieval process (Larson, 1996; Jones and Purves, 2008a; Purves and Jones, 2011). Leidner and Lieberman (2011) identify six main steps for the processing of textually-encoded spatial data (see Figure 2.2). This can be considered a prototypical model for those systems which deal with the relevance of geographic information.

First, documents in the collection have to be preprocessed. Metadata, formatting, layout, and other information which is not related to the content has to be pruned from the input document. This is usually restricted to separating the HTML[8] tags from a Web-page, or similar procedures to be applied to other digital document formats, but it can also include analog document scanning and optical character recognition procedures. The desired output is the textual part of the document, that will be analysed in the subsequent step.

The second step is called "geoparsing", and it is designed to recognise names of geographic entities within a text. The aim is to parse the text obtained from the first step and to recognise as many geographic name-entities as possible. A geographic name-entity is a series of one or more terms, which refer to a geographic entity. Although this is an every-day procedure for a human-being, it is quite difficult for a computer. The main problem is ambiguity (Overell, 2011; Buscaldi, 2011), i.e. a term can have different meanings. The term "Washington" may refer to George Washington, to the State of Washington, or to Washington D.C. – or to 43 other cities within the United Stats of America and 11 outside, according to Wikipedia[9]. Sometimes "Washington" is also used to refer to the government of the United Stats of America. This is a typical example

[8]HyperText Markup Language: `http://www.w3.org/TR/html`, last accessed July 2012.
[9]`http://en.wikipedia.org/wiki/Washington`, last accessed July 2012.

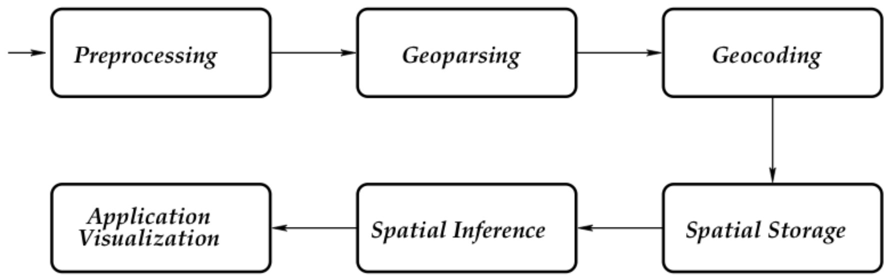


Figure 2.2: The six steps for the processing of textually-encoded spatial data identified by Leidner and Lieberman (2011).

of geographic vs. non-geographic ambiguity. Moreover, a structural ambiguity may also arise in a case such as “North Dakota”. It may be clear to a reader that it refers to the State of North Dakota, but for a computer it may not be straightforward whether to include “North” as a part of the geographic name-entity or not. Common approaches to this problem are: the use of gazetteers, such as GeoNames[10] (Manov et al., 2003); the use of symbolic rules depending on the language of the document (Schilder et al., 2004); and the use of machine learning algorithms (Curran et al., 2007).

The third step is called “geocoding” (Schlieder and Henrich, 2011; Schockaert, 2011), and it is designed to assign a spatial extent to the recognised geographic entities, which is commonly referred to as geographic footprint. This process may be easy for institutionally defined geographic name-entities, such as countries or cities. Nevertheless, not all countries have internationally agreed boundaries, and most big cities are also difficult to bound, as they might fade into suburbs, and merge with neighbouring towns. It is even harder to define the spatial extent of vague places such as a downtown or the Alps. For such cases, where no clearly defined boundaries exist, methods have been suggested to extract physical grounding (Jones et al., 2008; Straumann and Purves, 2008) or definitions of such vague places based on Web user-generated content (Schlieder and Matyas, 2009). This can be coupled with fuzzy systems, which are able to deal with non-crisp boundaries (Schockaert and De Cock, 2007; Bordogna and Psaila, 2012).

The fourth step is called “spatial indexing” (Leveling, 2011), and it is designed to efficiently structure, and store the extracted information, so that it can be easily retrieved. Traditional IR systems use an inverted index, where a term is linked to each document containing it. GIR systems require a more sophisticated index capable of organising the geocoded entities in a geographically meaningful manner. These indexing systems have to be able to support computation of geographic relationships among geocoded entities, such as proximity, inclusion, and exclusion.

The fifth step is called “spatial inference”, and it applies automated reasoning to the stored data. For example, a transitivity rule may be applied to geographic relationships of inclusion, encoded in the spatial index. A sixth step refers to any application,

[10] `http://www.geonames,org`, last accessed July 2012.

visualisation, or use of the stored data. In GIR, this step is commonly occupied by the relevance assessment and ranking, possibly coupled with a cartographic visualisation of the obtained rank (e.g., Purves et al., 2005).

The typical approach to relevance ranking in GIR consists of computing a score which is a weighted linear combination of a textual similarity score, and a geographic similarity score (Andrade and Silva, 2006). The textual similarity is commonly computed using standard IR methods, such as the Okapi BM25, or other td-idf methods. The geographic similarity score is computed based on topological relationships between the geographic footprint of the query and the geographic footprint of the document (Larson, 2011). If the footprints are geographic points, geographic similarity is commonly measured as the Euclidean distance on the Earth's surface. If one of the footprints is a geographic point, and the other is a polygon, the inclusion relationship is the most common choice. In fact, some systems allow the user to explicitly define a geographic footprint, e.g., by giving the user the ability to draw a polygon on a map. If both footprints are polygons, the most common metric are the area of overlap, and the Hausdorff distance (Larson and Frontiera, 2004). Alternative combination functions have also been suggested (Cai, 2002; Martins et al., 2005; Frontiera et al., 2008), such as product functions, and maximum score selection. Purves et al. (2007) found the non-distributed multi-dimensional scattered method proposed by Van Kreveld et al. (2005) to be to most suitable in the scope of the E.U. research project SPIRIT.

Cai (2011) stated that the methods described above still face several difficulties in finding their way to end-user services, and that most of the evaluations carried out on GIR systems had found little or no benefit in using the methods described above. Arguments have been put forward that this evidence is more due to the inadequacy of the evaluation methods, rather than to the deficiencies of the evaluated systems (Mandl, 2011). Moreover, Purves et al. (2007) found GIR methods to be profitable in the case of queries involving non-containment spatial relations (e.g., "15 miles north of Washington" or "near London"), and to be able to retrieve relevant documents which do not mention the place name specified in the query. The issues of system evaluation are discussed in Section 2.5.

2.3.1 Mobile information retrieval

GIR has been the first field to approach relevance as a dual concept, entailing a spatial facet along with the intrinsic information content (Larson, 1996). However, this is still not directly related to the location and the context of the user, but rather to a spatial extent implicitly or explicitly expressed by the user in a query. Moreover, in a typical GIR system, the documents do not have a spatio-temporal extent *per se*. They have a geographic footprint to which they refer, but they do not commonly have spatio-temporal accessibility constraints, and they are generally instantly available. Moreover the temporal information contained in the documents is seldom taken into account by GIR systems, although it influences the relevance of documents with respect to the user information need (Palacio et al., 2009, 2011). Hence, a user's location and potential mobility was not taken into account by GIR systems. This perspective has been drastically changed by the rise of mobile Internet access and mobile search engines.

This application of GIR to mobile usage was first investigated by Mountain (2005); Mountain and Macfarlane (2007) and named Mobile Information Retrieval (MIR). In this approach, the location of the user becomes a part of the query. Such a query is processed by a spatio-temporal IR algorithm, by means of four filters. The first is a spatial filter, which considers irrelevant any document whose footprint is farther away than a given distance threshold. A second filter is defined as temporal, but it does not involve just time but also its effect on space. This is unlike the criterion proposed by Bierig and Göker (2006), where the temporal availability of an entity or object is compared only to the information need "time", as past, present, or future. The temporal filter takes into account the user's mobility, where the two concepts of space and time are considered as independent.

In GIScience, this is related to the concept 'accessibility', which concerns the part of space that can be reached within a given amount of time. This concept can be used to assert whether a user is able to interact with a geographic entity, considering the travel time, the user needs to reach the entity, and respective spatial and temporal constraints (e.g., a user's scheduled appointment, or the opening hours of a shop). In this dissertation this concept will be referred to as 'spatio-temporal accessibility' or 'spatio-temporal proximity', in order to avoid confusion with the concept of 'information accessibility' as it is defined in IR (see Table 2.1). The space-time path and the space-time prism developed in time geography are the main models used to deal with 'spatio-temporal accessibility'. From an IR perspective, spatial, temporal, and spatio-temporal proximities are embedded in the concept of 'physical relevance' described by Reichenbacher (2005a, 2007), and in the 'horizon' concept described by Saracevic (1996) as a part of 'interpretational relevance'.

The third filter is based on the assumption that a user in a mobile environment is interested in what she can see in her immediate surroundings. The fourth filter developed by Mountain (2005) is the search-ahead filter. It is based on the assumption that users may be more interested in entities that are on their future path, rather than those that have been passed already. This implies some level of knowledge stored within the system about user's position along a path and the user destination, and the uses of time geography methods (see Section 2.2.2), or the usage of a prediction algorithm as developed by Mountain (2005).

Until very recently most of the studies in MIR have focused on rather small sets of geographically referenced documents. As the number of geographically referenced Web documents has increased in recent years, it became clear that more sophisticated approaches were needed in order to respond to mobile query efficiently. Within the field of very large databases systems, this has resulted in a growing interest in spatial Web objects (Wu et al., 2012). These objects are defined as pairs containing a Web document and a reference to a geographic location. The objective is to establish computationally efficient methods to retrieve the top-k objects (e.g., the top-10 objects) that satisfy a mobile user query considering both spatial proximity and textual relevance.

This line of research has undertaken a bottom-up approach, starting from current Web search engine indexing methods, which are enhanced to include the spatial distance between the referenced location and the user as a further criterion. Both Vaid et al. (2005) suggest to combine inverted index files and spatial index as an efficient indexing approach to this problem. Cong et al. (2009) use an r-tree, where each node includes an inverted file which provides textual indexing of those documents whose spatial footprint is contained in the sub-tree rooted at the node. The same approach has been then improved by Cao et al. (2010), who included a measure of the centrality of the referred location with respect to other referred location nearby. Venetis et al. (2011) built on those previous works, incorporating as a new criterion the frequency with which each place has been mentioned into direction queries on Google Maps.

2.3.2 Mobile cartography

GIScience and IR are not the only disciplines which have encountered considerable changes due to the "mobile revolution". The discipline of cartography has also been influenced by this progress at the beginning of the millennium when recently-developed Web maps have made the leap from desktop computers to mobile phones (Meng, 2005). In a few years, cartography moved from the first web maps to the opportunities offered by Internet-enabled location-aware mobile phones, able to dynamically load information from the Internet, based on the location of the user. It has become possible to produce adaptive geographic information visualisations on-the-fly on mobile phones. In this context, the new field of mobile cartography has been put forward (Reichenbacher, 2004).

The main goal of this new field is to offer visualisations of geographic information with the greatest possible relevance to a mobile user. Within this framework, the adaptation of the map goes hand in hand with the relevance of the information being displayed, as suggested by Reichenbacher (2005a,b, 2007, 2008). The adaptation of the map should be commensurate to the relevance of the entity. This concept is represented in Figure 2.3, where the first adaptation step reduces the amount of information, and the second step modifies how the objects are being visualised, based on their relevance. Figure 2.4 gives an example of such a process, where the map on the right-hand side is the result of adapting the original map in the left-hand side to the geographic information need of a user searching for a cafe with an ATM nearby. In this example, the applied procedure fades the base topographic layer and the less relevant entities (i.e., the cafes without an ATM nearby), in order to make the more relevant entities (i.e., the cafes with an ATM nearby) more visible.

In mobile cartography, the user context has long been recognised as a key component in adapting mobile maps to the user's needs. The use of context and activity theory applied to the assessment of geographic objects' relevance in the field of mobile adaptive

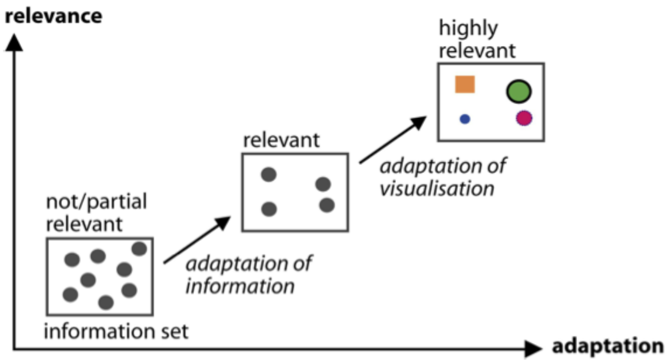


Figure 2.3: Relationship between map adaptation and information relevance, as suggested by Reichenbacher (2007).

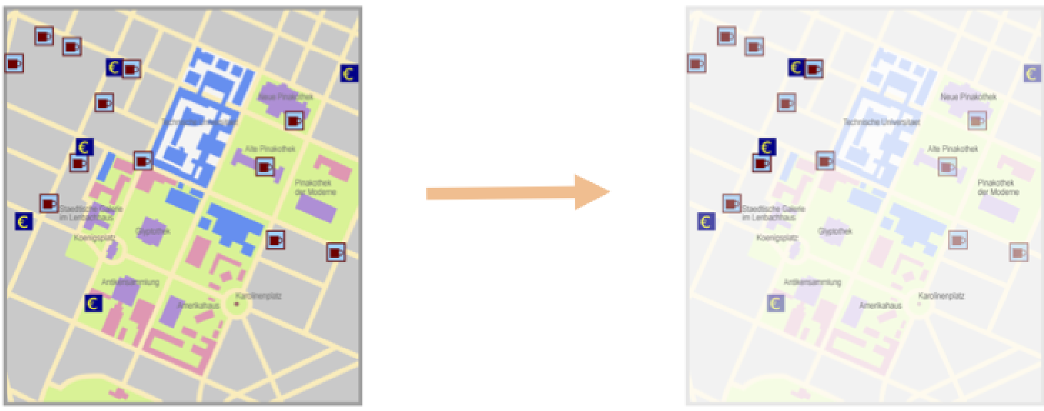


Figure 2.4: Example of mobile map adaptation Swienty and Reichenbacher (2006): the map on the right-hand side is the result of adapting the original map in the left-hand side to the geographic information need of a user searching for a cafe with an ATM nearby.

Figure 2.5: Screenshots of the *Hike & Bike Map* application for Saxon Switzerland, developed by Hauthal and Burghardt (2012).

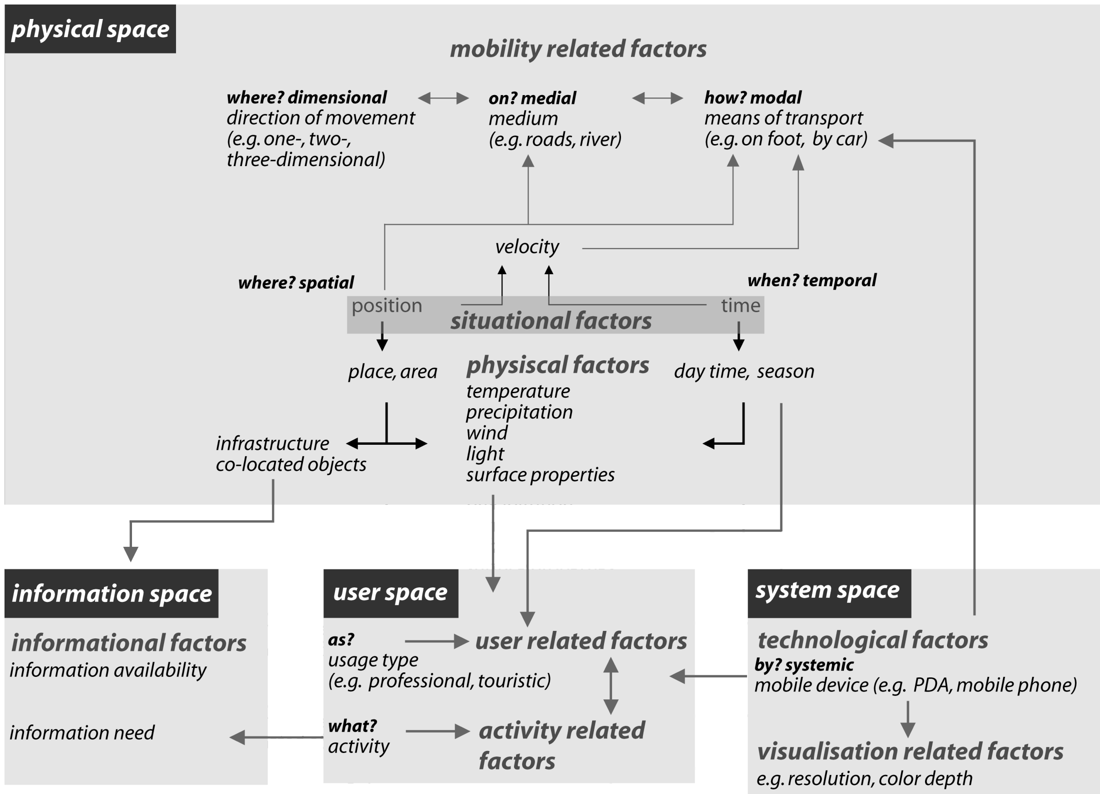


Figure 2.6: Mobile context factors and their interrelationships, as proposed by Reichenbacher (2008).

maps has been discussed by Reichenbacher (2005b) and Swienty et al. (2008). Figure 2.6 presents an overview of context factors in mobile cartography, and describes how the user's activity is part of the context, and how it interacts with the other factors. The adaptation of the mobile geographic information visualisation to the context through the concept of relevance can significantly improve the usability of the geographic information content on mobile devices (see Figure 2.3).

Most LBS are map-based applications, thus mobile cartography (Reichenbacher, 2001) brings a crucial contribution to that field of research. Typical examples are mobile guide applications, which provide touristic information to a mobile user in the form of a LBS. An example is the *Hike & Bike Map* developed by Hauthal and Burghardt (2012) that can suggest a list of possible trips in the user's vicinity (see Figure 2.5). When a trip is selected, related points of interest (POIs) are shown on a map. In this particular application, each trip has a pre-defined set of POIs to be displayed. This type of service would require a relevance assessment as soon as there are too many POIs to be displayed on a mobile map. This might be even more critical, if such a service is connected to large POIs datasets, such as those available from OpenStreetMap[11] or private companies. Similarly, Huang and Gartner (2012) present a POI recommendation system for the Vienna Zoo (Austria). The system suggests visitors which POIs to visit. The recom-

[11] `http://www.openstreetmap.org/`, last accessed April 2012.

mendation is based on collaborative filtering. A preliminary set of context parameters (e.g., age and time limit) is integrated with a further set of parameters identified through trajectory mining, where the behaviour of the current user is compared to past visitors of the zoo, searching for similar trajectories.

In fact, data mining techniques can be resourceful tools for geographic knowledge discovery. The mined information about user's mobility and geographic environment can be included as part of the context of usage in mobile applications. Reichenbacher (2004) also mentioned that clusters and co-location of different types of entities in the geographic environment can be seen as part of the context to be taken into account. The subject of geographic data mining and information analysis is discussed later in Section 2.4.

2.3.3 Geographic relevance

As argued above, location by itself is not sufficient for understanding context for mobile geographic information services. The next generation of LBS must look at more than just location in order to understand the user's need for geographic information. In order to face this challenge, a new concept of relevance has been outlined (Zipf and Richter, 2002; Coppola et al., 2004; Reichenbacher, 2005a; Raper et al., 2007b).

The fundamental idea is that the concept of relevance should not only be applied to documents, but also to physical entities in the geographic space, from a user's perspective, in a given spatio-temporal situation. To label this idea, Raper (2007) coined the term "geographic relevance" (GR). This can be defined as a quality of an entity in geographic space or its representation in an information system (i.e. an object, document, or image), expressed as the relation between an entity or its representation and the actual context of using the representation (Reichenbacher et al., 2009). An in-depth discussion on how the concept of GR has been developed in the last few years within the scope of the GeoRel project is presented in Chapter 3.

GR does not aim to assess the relevance of documents, thus it may not be considered within the boundaries of IR as it is commonly understood, as well as within GIR and MIR. The latter focus on the geographic information content of documents, whereas GR focuses on geographic entities as abstractions of real-world objects. This concept is similar to relevance as it is understood in the field of spatial Web objects, however, GR does not assume a direct link between geographic entities and Web pages. Moreover, the line of research focusing on spatial Web objects undertakes a bottom-up approach and it is more focused on the efficiency of the process, rather than on the effectiveness of the outcome.

Finally, GR is not concerned with the fitness for use of whole spatial data sets for particular GIS applications. Hence, this dissertation does not fall into the fields of spatial data quality and fitness for use of spatial data sets (Chrisman, 1984; Guptill et al., 1995; Morrison, 1995; De Bruin et al., 2001). Although the quality of the spatial data set affects the geographic information served to a mobile user, GR is concerned

Previous definitions of geographic relevance

"Geographic relevance [can be] defined as a relation between a geographic information need and the spatio-temporal expression of the geographic information objects needed to satisfy it in order to take some action. " (Raper, 2007, p. 846)

"Geographic relevance is a quality of an entity in geographic space or its representation, i.e. an object, document, or image. This quality is expressed as the relation between the entity or its representation and the actual context of using the representation. The elements in the geographic representation can be discrete objects, properties of the objects, relations between objects, or the structure of objects. The main dimensions of the usage context are theme, space, time, intention, and knowledge state." (Reichenbacher et al., 2009, p. 1)

with the elements within the dataset, not the whole data set. The general objective of relevance assessment is to understand which elements in the dataset are relevant, not the appropriateness of the whole collection. Although, some of the data quality elements may be considered to be important dimensions of relevance, they are not necessarily sufficient to qualify an object as relevant.

2.4 Geographic data mining and information analysis

The reason for using relevance concepts in mobile information services is to prevent information overload, and thus fulfil the user's geographic information needs. However, the increasing amount of geographic data available gives us new opportunities as well. The automated analysis of raw data can produce new information to be used in the information retrieval process. In particular, this wealth of information can be exploited to analyse first- and second-order effects in the distribution of the entities in the geographic space (e.g., O'Sullivan and Unwin, 2003), which can then influence the GR of an entity as part of the context in which that entity exists – as further discussed in Chapter 3.

As text data mining is most useful for document IR, geographic data mining can be an important resource for mobile geographic services, such as LBS and mobile cartography, as has been suggested by Reichenbacher (2004). LBS and mobile cartography services, thus far, have focused their attention on single entities, with little or no attention being paid to the surroundings of those entities. However, entities do not exist as independent objects, but rather are embedded within a specific geographic context. Geographic information analysis (O'Sullivan and Unwin, 2003) and data mining techniques (Miller and Han, 2001, 2009) can be fundamental tools for understanding complex geographic phenomena that involve not only a single geographic entity, but also its relationships to its surroundings and neighbouring entities in a larger relational context.

Spatial clusters are among the most evident and prominent first-order effects, which can be understood as part of the geographic context of an entity. Several spatial clus-

Geographic data mining and information analysis

"Geographic data mining involves the application of computational tools to reveal interesting patterns in objects and events distributed in geographic space and across time (in very large geographic datasets)*."* (Miller and Han, 2001, p. 16)

"A working definition of geographic information analysis is that it is concerned with investigating the patterns that arise as a result of processes that may be operating in space. Techniques and methods to enable the representation, description, measurement, comparison, and generation of spatial patterns are central to the study of geographic information analysis." (O'Sullivan and Unwin, 2003, p. 3)

tering methods have been developed from statistical cluster analysis over the years, including both partitioning methods, and hierarchical methods (Han et al., 2001). The main disadvantage of partitioning methods inherited from statistics is that they can only find spherical-shaped clusters, whereas geographic clusters are often non-spherical, since they evolve in geographical space, in which the distribution of entities is constrained by geographic features (e.g., lakes, rivers, mountains, and streets) (Sander et al., 1998). Hierarchical methods perform better in this task. Other clustering methods define clusters as areas with relatively high object density. These are referred to as density-based methods, such as the method DBSCAN proposed in (Ester et al., 1996), which can effectively discover non-spherical clusters.

While spatial clusters describe the surroundings of an entity in terms of other similar entities nearby, the concept of co-location describes a second-order effect between an entity and surrounding entities belonging to other categories. Two instances of different categories are said to be co-located if they are within a defined threshold distance from each other. Therefore, given a dataset of entities and a category among those present in the dataset, a co-location rule defines the categories whose instances are frequently found co-located with an entity belonging to the first category. This knowledge is an important piece of information in understanding the geographic surroundings of entities. In fact, it complements the information obtained from the cluster analysis, integrating information about spatial relationships with entities of the same category with information about spatial relationships with entities of the other categories. Several techniques have been proposed (Huang et al., 2004), which mine spatial databases to find such co-location rules. More sophisticated methods have been developed, which can also discover negative co-location rules (Jiang et al., 2010). These second type of roles capture the fact that certain categories of entities are usually absent in the neighbourhood of another given category.

Both the first- and second-order effects mentioned above are still limited to the spatial dimension. A different type of analysis has to be carried out, if one is interested in discovering rules that describe relationships among entities along the spatial, temporal, and attribute-related axes. For this purpose, techniques have been designed to mine as-

First- and second-order effects

"First-order *spatial variation occurs when observations across a study region vary from place to place due to changes in the underlying properties of the local environment. For example, the incidence of crime might vary spatially simply because of variations in the population density, such that rates increase near the center of a large city. In contrast,* second-order *variation is due to local interaction effects between observations, for example, the existence of crime in an area making it more likely that there will be crimes surrounding that area, perhaps in the shape of local hot-spots in the vicinity of bars and clubs, or near street drug markets. In practice, it is difficult to distinguish between first- and second-order effects."* (O'Sullivan and Unwin, 2003, p. 29)

sociation relationships among spatial and non-spatial predicates, discovering association rules among entities and their properties (Koperski and Han, 1995).

Understanding user mobility as a spatio-temporal process can provide additional information about the usage context of a mobile application. Current research in geographic data mining is also focusing on knowledge discovery from movement databases (Giannotti et al., 2008; Dodge et al., 2008; Miller and Han, 2009; Dodge et al., 2012). Techniques, such as trajectory classification, trajectory clustering, and movement data similarity analysis are used for mining common movement patterns among different users. This can be very useful for mobile application development. The recognition of particular patterns of movement (e.g., 'meet' or 'moving cluster') can provide important insights into the user's context (e.g., group activities). The similarity between the current user's trajectory and other users' trajectories could make it possible to predict a user's future spatial behaviour, similarly to what has been suggested by Huang and Gartner (2012). The same techniques can be used to further refine the probability distribution employed in the search-ahead filter suggested by Mountain and Macfarlane (2007).

2.5 System evaluation

System evaluation has always been one of the main topics of IR, involving many areas including information-seeking behaviour analysis, usability of the interface, and computational complexity analysis of the systems in terms of storage and time (Sanderson, 2010). Since the earliest stages of IR, the evaluation of IR systems has been performed by means of a test collection and an evaluation measure, which simulate a user of a searching systems in an operational setting. A test collection is composed of a set of documents, a set of topics, and a set of judgements reporting the relevance of each document with respect to each topic. The judgements are commonly performed by human analysts. The relevance assessments computed by the system are compared with the judgments specified in the test collection: the more similar the two are, according to the selected measure, the higher the system effectiveness.

It is possible to compare the effectiveness of two or more IR systems, by using an agreed test collection and evaluation measure. The same procedure can also be used to study the effectiveness of an IR system with different configurations. Starting in 1992, the Text Retrieval Conference (TREC) has been the largest evaluation series in IR (Voorhees et al., 2005). Large evaluation series such as TREC, the NII Test Collection for IR Systems (NTCIR) (Kageura et al., 1997; Kando et al., 1999) in East Asia, and the Cross-Language Information Retrieval (CLEF) (Peters and Braschler, 2001) in Europe made it possible to compare various IR systems, on the basis of a standard procedure and a shared collection of documents. The main critiques to this type of evaluation concern the assumed independence of relevance of the documents in a collection, and the subjectiveness of relevance judgments (Manning et al., 2008).

In GIR, the GeoCLEF (Gey et al., 2007; Mandl et al., 2008, 2009) evaluation series has been organised from 2006 to 2008 within the CLEF series. The results collected by GeoCLEF show that GIR systems could at best perform as naïve systems based on standard IR procedures. A different approach was taken with the GikiCLEF (Santos and Cabral, 2009, 2010). In that series, the systems were asked to first generate an answer to a geographic question, and then use this answer for the document retrieval. These experiments have shown a clear difference between questions conveying a strong geographic scope and questions conveying a loose geographic scope. GIR systems are more effective than IR systems in dealing with the first type of questions, as discussed in Section 2.3, see also (Purves et al., 2007).

At the time of writing, the first edition of the Contextual Suggestion Track[12] is running as part of TREC 2012[13]. This evaluation track focuses on recommending local businesses, based on the user's profile and the context. The latter is defined as the city where the user is, the time of day, the day of the week, and the season. Despite the vague definition of both the spatial and temporal information, this is the first example of an evaluation track involving the relevance assessment of geographic entities[14].

2.5.1 Crowdsourcing

Large evaluation campaigns such as TREC are not always affordable for small interest groups, focused on subfields of IR. In the last few years, crowdsourcing (Marsden, 2009; Eickhoff, 2011; Yuen et al., 2011) emerged as an alternative path to IR evaluation (Alonso et al., 2008; Alonso and Mizzaro, 2009; Alonso and Baeza-Yates, 2011).

Crowdsourcing can be defined as the mechanisms and associated methodologies for scaling and directing crowd activities to achieve some goal(s) enabled by Internet connectivity. Crowdsourcing platforms such as Amazon Mechanical Turk[15] are Internet services that give developers the tools to create and submit tasks to a wide audience

[12] `https://sites.google.com/site/treccontext/`, last accessed September 2012.

[13] `http://trec.nist.gov/pubs/call2012.html`, last accessed September 2012.

[14] Nevertheless, the vagueness of context information and the small amount of data involved are not compatible with the GR assessment model presented in this dissertation, making it impossible to use it as evaluation procedure of the presented model.

[15] `https://www.mturk.com`, last accessed May 2012.

of registered users, through an Application Programming Interface (API). The obtained answers can be incorporated into software applications, and the manner of communication resembles a common remote service. The response times are commonly rather short, due to the large number of members collected by these services.

This results in a very fast and relatively cheap way to place questionnaires, or run any kind of user experiment, which can be incorporated in a Web page and run without particular equipment. The participants can be assumed to be competent computer users or at least familiar with the Web environment. Studies performed on the demographics of Amazon Mechanical Turk (Ipeirotis, 2008, 2010; Ross et al., 2010) showed that in 2008 most of the participants were residents of the Unites States of America, whereas participants from India accounted for almost half of the population in early 2010. The same demographics reported that about two thirds of the participants from India have at least a Bachelor's degree, and about one third of the participant from India declared that the money gained through Amazon Mechanical Turk is (at least sometimes) "necessary to make basic ends meet". This raises ethical concerns, as discussed by Felstiner (2010). At the same time, it offers quite a unique opportunity to perform experiments with such a diverse set of subjects (Mason and Suri, 2011).

From an experimenter's point of view, a key issue is the quality of the obtained answers (Marsden, 2009). There is no guarantee that participants will carry out the tasks in a reliable and foreseen manner. Malicious workers have been reported by almost every study made on crowdsourcing platforms, especially in the earliest years. Nowadays, most platforms offer quality check systems. The most common approach is to check the evaluation that previous experimenters have assigned to a given participant. Several studies (Eickhoff and de Vries, 2011; Harris, 2011) reported that the quality of the results can be improved by using open questions, screening questions, or more complex tasks, which discourage malicious workers. A training module prior to complex tasks is suggested, to guide the worker through some warm-up questions and avoid misunderstandings.

Crowdsourcing has been applied to particular IR tasks, such as video annotations (Soleymani and Larson, 2010), music similarity assessment (Urbano et al., 2010), and news search (McCreadie et al., 2010). Despite the issues mentioned above, these platforms offer an opportunity for collecting human judgments for a small scale evaluation dataset, with a very high cost and time efficiency.

2.6 Implications

In this chapter, I reviewed different fields of research which investigate the concepts of relevance, context, and mobility (see Sections 2.1 and 2.2). I also summarised the studies focusing on the relationships between those three concepts, and their application to the development of mobile information services (see Section 2.3). Thereafter, I outlined how mobile information services can be strongly affected by information overload and related issues, and how the current limitations of these services led to the proposition of the concept of GR (see Section 2.3.3). Finally, I briefly presented the key tools (see

Sections 2.4 and 2.5) used to develop and evaluate a novel GR assessment method in the presented research.

The concept of GR aims to integrate and encapsulate ideas and concepts from different academic communities working on the development of mobile information services. However, GR still lacks a development framework, a list of relevance criteria, and a formal method for the numerical assessment of relevance scores. This dissertation aims to fill that gap.

It is possible that the information overload problem under investigation could be solved by applying models and criteria developed in the field of IR to the available information about the geographic entities. This would mean that GR is not a new concept, but rather a reformulation of the concept of relevance commonly used in IR, applied to different underlying data. It is therefore necessary to further investigate the conceptual basis of GR, in its differences from the concept of relevance commonly used in IR, and in its geographical aspects beyond the sole spatial proximity criterion. Further criteria related to the user's mobility and the geographic environment may play an important role. These topics will be discussed in Chapters 3 and 4.

On the one hand, if GR is just a reformulation of the concept of relevance commonly used in IR, the assessment methods used in document-based IR could be directly applied to information available on a given set of geographic entities without further adaptation. On the other hand, if GR is a new concept of relevance and different criteria have to be taken into account, a new assessment model would be needed, as well as new formal definitions to numerically estimate GR with respect to those criteria. These topics will be discussed in Chapters 5, 6, and 7.

Chapter 3

Geographic relevance

As suggested by Schamber et al. (1990), relevance is a multidimensional, dynamic, and complex concept. Since the 1960s, information science has invested great effort into understanding its nature, and many definitions of relevance has been proposed. Nevertheless, the same nature of such concept implies the existence of different types of relevance (e.g., Borlund, 2003). In this chapter, I present the concept of Geographic Relevance (GR) and the conceptual framework of this dissertation. The mobile revolution (as reviewed in the previous two chapters) has made it necessary to apply and adapt concepts and methods developed in IR to mobile cartography and LBS. The key tool in this process is the concept of GR.

Section 3.1 provides a definition of GR, a description of how this concept is derived from previous concepts of relevance, and a discussion on how it relates to these prior definitions. Section 3.2 discusses how different models of space can influence the conceptualisation of GR. Finally, a conceptual model is proposed in Section 3.3, which presents the elements identified as key to describe the user context, geographic entities, and the surrounding environment.

3.1 Definition and derivation

As discussed in Section 2.1, relevance is still an ill-defined concept. In IR, relevance is commonly defined as a relationship between a user need and an entity. An extensive survey on the development of the concept of relevance in IR is presented by Mizzaro (1997a). GR is the relation between an entity in geographic space, as represented in an information system, and the geographic information need of a user, as expressed in the language of the system, in the context of using the information system. This definition of GR is derived from the definition proposed by Raper (2007), who based his definition of GR on the widely accepted definition of situational relevance, as proposed by Wilson (1973).

Situational relevance combines two relevance concepts. The first is logical relevance, as developed by Cooper (1971) more than forty years ago. Cooper defined logic relevance as '*a relationship holding between pieces of stored information on the one hand and user's* *information needs formulated as information need representations on the other hand*' (Cooper, 1971, p. 22). Cooper depicted a user's question (i.e., his conceptualisation of the user's information need) as the sets of its possible answers. If at least one of those answers logically follows a piece of stored information, the latter is defined as logically relevant. This type of relevance is directly derived from deductive logic. The second concept entailed in situational relevance is evidential relevance. This is an inductive type of relevance, which conveys '*the notion of degree of confirmation, or probability, of conclusions in relation to given promises*' (Wilson, 1973, p. 460), which is an intrinsic factor of human knowledge and reasoning, and calls for the usage of plausible or probabilistic inference in IR. Hence, situational relevance is the relevance of a piece of information to the user's situation as she sees it, not as it "really" is. The logical acceptation of relevance is the same embodied in logic relevance, but its clarity is "infected" by the indeterminacy of evidential relevance.

This is just one of the various conceptualisations of relevance which has been proposed since the 1960s in IR. To bring some order to the proposed concepts and definitions, Mizzaro (1998) suggested four dimensions along which various types of relevance can be characterised. These dimensions are 'information resources' (i.e., what the user is searching for), 'user problem', 'time', and 'components'. In this model, relevance can deal with the information resources at three levels of abstraction: a document containing information; a surrogate information of the documentation (e.g., the title and keywords); and the information received by the user, as she perceives it. The user problem has four levels of abstraction: her real information need (RIN); the information need as she perceived it (PIN); the information need as she can express it in a natural language; and the query as it is expressed in the system language. The dimension 'time' refers to the steps in which the information flows with the interaction from the moment in which the user's real information need arises to the moment in which it is satisfied. The dimension 'components' lists the aspects that compose the first two dimensions: the topic the user is interested in; the task or activity she aims to perform; and the context in which everything happens.

In this framework, Wilson's situational relevance (Wilson, 1973) can be defined as concerning the information as the user receives it, and the user's information need as she perceives it – Wilson's definition of situational relevance does not address the dimensions 'time' and 'components'. It is evident that this framework does not account for the concept of GR, because GR does not deal with the relevance of documents.

Mizzaro's framework was further developed by Coppola et al. (2004) when defining w-relevance. The authors use the "w" to refer to wireless relevance, but also double-relevance, world-relevance, and double-user-relevance (see Coppola et al., 2004, p. 7). This concept partially anticipated the notion of GR, in the sense of relevance addressing real/physical world entities and not documents. In this new framework, relevance can deal with the resources at four levels of abstraction: the actual entity (or thing); a document describing, representing, or referring to the entity; a surrogate information of the entity or its documentation; and, the information received by the user, as she perceives it.

This further level of abstraction is a key change, because it reveals a whole new layer of relevance concepts, beyond the classic horizon of document-based IR. These relevance concepts address the physical objects that are at the core of current e-commerce systems, Recommender Systems (RS), and LBS, instead of digital documents that describe and refer to physical objects. For instance, if a user is searching for a restaurant for lunch, a webpage containing information about local restaurants can be considered relevant by a document-based IR system, although some of the restaurants described in the webpage might not be considered relevant by a LBS (for example, because they are closed). However, three main issues should be addressed for this framework to be able to deal with spatial and context-aware definitions of relevance, as GR aims to. The first relates to a user's mobility, the second to the perception of the context, and the third to the dimension 'components'.

Mizzaro (1998) describes how the relevance of a document (or any other level of information resource abstraction) to a query (or any other level of information need abstraction) changes over time, as the user acquire information, and thus her information need changes. This fact is captured in the framework by means of the dimension 'time', which has been suggested accounting for an information seeking scenario, where a user is supposed to be sitting at a desk. This "static" assumption does not hold anymore, as users can now search for information on their mobile devices. In fact, a mobile user can "trade" time for space in order to generate mobility (as discussed in time geography, see e.g., Miller and Bridwell, 2009). One can spend time travelling, obtaining in return the movement from a location to another (i.e., mobility). A user's location can change as time passes by from the moment the RIN arises until it is met, and location can be a fundamental factor of relevance (Reichenbacher and De Sabbata, 2011). Thus, a piece of information which is relevant for the location where the user is when she perceives her information need may no longer be relevant for the location where she is when she obtains the retrieved information, because of the change in location.

Hence, a relevance dimension concerning time can not ignore space, as these two factors are interdependent. Therefore, I suggest to replace the dimension 'time' with a dimension concerning space-time. This new dimension 'space-time' describes how the information flows with the interaction, from the point in space and time in which the user is at the moment her RIN arises, to the point in space and time in which the user is at the moment the information need is satisfied. It has to be noted that the dimension 'space-time' is used to define which user position in space and time is considered in assessing relevance, and it does not relate to the information need or information content. For instance, the GIR systems discussed by Palacio et al. (2010) consider the spatial and temporal content of documents (i.e., as part of the topic the user is interested in, within the information resource dimension), but not the spatial and temporal context in which the information seeking happens (i.e., the 'space-time' dimension).

The second issue mentioned above concerns the definition of context. This is becoming a key factor in current definitions of relevance. Context is a vague term, which is used to refer to very different components of relevance, from a small set of raw input from a device's sensors to an abstract conceptualisation of the world around a user. As the RIN (i.e., the real information need) is different from the query received by the system, the real world is different from the world perceived by the system. Similar notions in GIScience are based on conceptual models such as the one proposed by Peuquet (1984, 2002), where '*levels progress from reality, through the abstract, user-oriented information structure, to the concrete, machine-oriented storage structure*' (Peuquet, 1984, p. 69). Hence, a new dimension should be considered to handle the representation of the world in which the relevance relation between the user's information need and the entity takes place (i.e., the context, both informational and physical). A relevance concept can be described as dealing with reality at the following levels of abstraction:

- *real world*;
- *documented world*, that is the world as it is recorded by the human knowledge in any form of stored information;
- *perceived world*, that is, the world as known and perceived by a user;
- *system world*, that is, the world as the systems knows it (i.e., the data available both from resources of the system and sensors of the user's devices).

This is a partially ordered set. The real world is (ideally) the most complete understanding of reality. It is reasonable to picture the documented world as a more complete approximation of the real world than the perceived world, because there is much information documented that remains unknown to a user. It is as well reasonable to imagine that a user can perceive information about the world which is not in the documented world, and that can be just partially documented in real-time by its device and sent to the system. The system's world is a subset of the documented world.

In addition to the two issues discussed above, I advance an argument for the dimension 'components' to be an attribute set more than an actual dimension. The distinction between this and the other dimensions is already evident in the definition given by Mizzaro (1998). The components are used to qualify the abstraction levels defined in the other dimensions (excluding the dimension 'time'), rather than to define the steps of an independent dimension. Moreover, I suggest to extend the list of components in order to explicitly include the user's preferences, social context, and mobility, which are among the most frequently used facets of relevance. Therefore, the components of relevance included in the framework considered in this dissertation are:

- *topic* refers to the subject the user is interested in;
- *activity* (or *task*) refers to how the retrieved information is concerned with the activity or task that the user will perform, how the activity aroused the information need, and how the user will employ the information in pursuing the activity;

- *preferences* refers to the user's preferences related to the topic or the activity considered (e.g., Rashid et al., 2002);
- *social* refers to the social context that can influence the relevance of an entity, such as the popularity within a community or among user's connections (e.g., Mizzaro and Vassena, 2011);
- *mobility* refers to the spatio-temporal availability of the entity (e.g., location and opening hours), the spatio-temporal situation of the user (e.g., location, time schedule, and mode of transportation), available mobility infrastructure, and topological structure (e.g., Mountain and Macfarlane, 2007);
- *context* is everything not pertaining to the previous components, including knowledge about the physical surroundings (e.g., light level or other geographic entities in the surroundings) and informational surroundings (e.g., information contextual to the topic or activity).

Using this framework, relevance as it is commonly implemented in GIR systems can be defined as being concerned with documents describing geographic entities and the user's query (see Table 3.1). These elements are described using the component topic. MIR systems usually include the mobility component. Relevance as it is commonly implemented in LBS can be defined as being concerned with a surrogate information about geographic entities and the user's query, at the spatio-temporal point at which it is submitted to the system. The elements are described using the topic and mobility components, although this last, in most cases, is restricted to user's and entities' locations in a simple geometric representation of space. Most RS take into account the topic, social, and preferences components. More advanced LBS and RS systems account for the activity and context component, where the latter is usually developed in its informational aspect.

As defined by Raper (2007), GR can be seen as an extension of Wilson's situational relevance. Assuming Raper's definition, GR is concerned with the information as the user receives it, and the user's information need as she perceives it, at the spatio-temporal point at which the information is perceived by the user. GR is concerned with the perceived world if the 'egocentric' attention is assumed, and it is concerned with the documented world if the 'allocentric' attention is assumed (Raper, 2007, p. 846). Finally, the elements involved in the GR relationship would be described by the components topic, activity, mobility, and context. This is a quite abstract definition of GR, but Raper (2007) also proposed a preliminary concrete model of GR based on the projection of a user's need on the world representation (referred to as user attention) and the entity footprint (referred to as influence of the geographic information object) onto a triangular irregular network representing the geographical space. The intersection of the two is the quantification of GR.

Aiming to define a computational method to assess GR, in this dissertation I employ a more pragmatic definition of GR. This definition is closer to the concept of relevance as

Table 3.1: Different understandings of relevance characterised in the proposed framework.

Definition	Dimensions of relevance				Components
	Resource	Need	World	Space-time	
Wilson Wilson (1973)	information	perceived information need			
Raper Raper (2007)	entity	perceived information need	perceived (egocentric) or documented (allocentric)	information perception	(topic, activity, mobility, context)
GIR	surrogate (document)	query			(topic, activity)
MIR	surrogate (document)	query		query submission	(topic, activity, mobility)
RS	surrogate (entity)	query			(topic, activity, preferences, social)
LBS	surrogate (entity)	query		query submission	(topic, activity, mobility)
GR	surrogate (entity)	query	system	query submission	(topic, activity, preferences, social, mobility, context)

defined by Saracevic et al. (1988); Saracevic and Kantor (1988a,b), which was described in Mizzaro's (1998) framework as dealing with the surrogate of the information resource, the user's request, and the components topic, task, and context.

In the scope of this dissertation, I define GR as concerning the surrogate information as it is available for the information system, and the user's information need as it is formulated in the query language, at the spatio-temporal location at which the user submits the query to the system. GR is concerned with the system's world, including the information already available to the system and the incoming information from the user's device. Finally, the elements involved in the GR relationship are described by the components topic, activity, preferences, social, mobility, and context.

One the one hand, it is clear that the main difference between GR and situational relevance is the focus on the actual entity, not on a document referring to that entity. This affects the difference between the more pragmatic definition of GR given above and relevance as defined in GIR and MIR (which are pragmatic derivations of situational relevance). The surrogate information taken into account in assessing GR is a surrogate of the entity (i.e., its representation within the information system), not a surrogate of a document as in IR, GIR, and MIR. On the other hand, the purpose of GR is to embody the concepts of relevance which are developing in the fields of mobile cartography, LBS, spatial Web objects, and RS. GR acknowledges a user's activity, the informational context in which an information seeking process happens, and aims to further develop the understanding of the geographical context in which the geographic entities are placed.

This is the understanding of relevance to be analysed and modelled in the following sections. In Chapters 4 and 5, I will investigate the criteria needed in order to compute this relevance, and propose a computational model for the GR assessment. Although for the sake of completeness all the identified components (i.e., topic, activity, preferences, social, mobility, and context) have been included in the definition proposed above, not all of them can be further investigated in detail within this thesis – due to the limited temporal scope of the project here described. Future developments in those directions are discussed in Chapter 9.

Wilson's (1973) situational relevance was defined in contrast to the notion of psychological relevance. Situational relevance is not concerned with whether and how the user perceived the relevance of a piece of information, nor with whether their view of the world has changed or failed to change after having received that information. The same holds for GR. In a first step, given the user context and a geographic entity, the GR of such entity can be assessed. In a second step, the assessed GR can be represented (Crease and Reichenbacher, 2011), for example, on a map on a user's mobile phone. The psychological relevance of the geographic entity is the result of the user perceiving the represented GR. Dealing with this psychological relevance is neither the aim of IR, nor of this dissertation. In order to establish a computational method for the assessment of GR, given an individual stock of information, and a structure of values or preferences, relevance is here considered in its logical aspects, and not in its psychological ones (as in Wilson's perspective).

3.2 The role of space

The 'mobility' and 'context' components included in the framework prosed in this work play an important role in defining the relevance of physical and geographic entities. This is the case for spatially- and context-aware concepts of relevance such as GR (De Sabbata and Reichenbacher, 2012). At the same time, these components are strongly influenced by the conception of space taken into account by the underlying information system.

Located at the intersection between geographic information systems and mobile services (Brimicombe, 2008), LBS usually take into account a simple model of space, where entities are modelled as points, lines, or polygons in an empty bi-dimensional Euclidean space. However, the mobility and context components of relevance require more sophisticated fundamental geographic concepts (Golledge, 2002), such as proximities, spatial relationships, geographic associations (e.g., neighbourhood relationships, clusters, co-locations, etc.), as well as spatio–temporal constraints (e.g., accessibility). For instance,

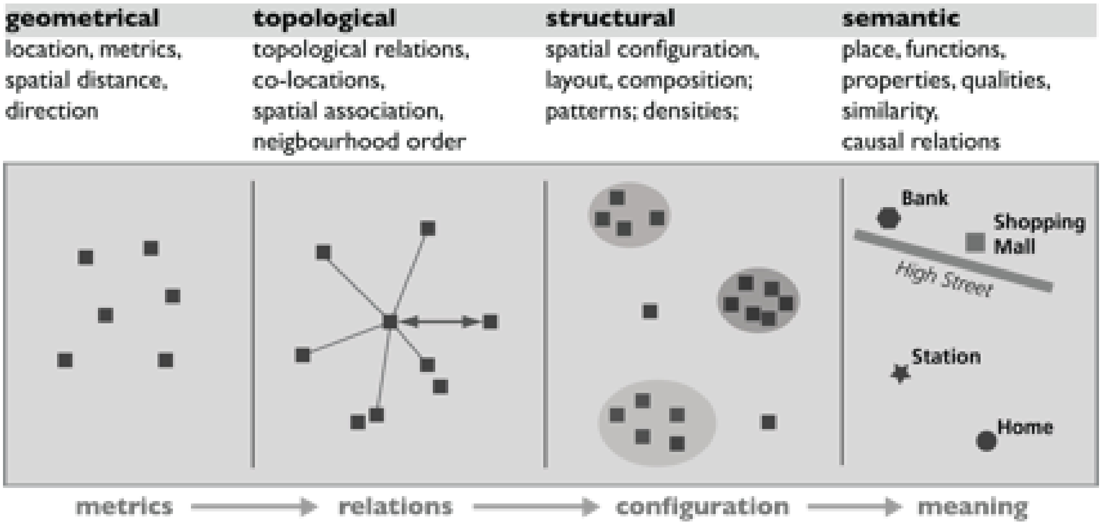


Figure 3.1: Different concepts of location and space.

a pure geometrical model of space is not sufficient to study and understand GR, which must be analysed with different models or conceptions of space. Raper (2007) distinguishes two geographic models of space. A "geo-representation" (i.e., a space modelled with geometric entities), and a "geo-context" (i.e., a mental constructs, such as places or landmarks). According to Reichenbacher (2009), four different models of space can be defined (see Figure 3.1), which are characterised by increasing complexity, richness, and expressive power. They include metrics, relations, configuration, and meaning.

The geometrical model of space uses geometrical references to space, including locations, areas, distances, and directions. This model is sufficient for determining simple proximities in space as in a typical LBS use case. Often one is not only interested in metrics, but also wishes to know about the relation of an entity to other entities. For capturing such spatial relations and associations, the geometrical model has to be extended to a topological model of space. The latter additionally captures spatial relations and associations required for assessing the relevance based on connectivity and accessibility in a network. Accessible space denotes the parts of space that can be reached within a specific time. For instance, parts of space not accessible within a reasonable time may be considered as seldom relevant for a user in a mobile information seeking scenario.

The structural model of space extends the geometrical and topological model by encompassing spatial configurations, layouts, compositions and spatial patterns. Examples are dispersion, clusters of entities, or densities, and include spatial primitives such as those proposed by Lynch (1960). For instance, an area where the density of a given category of entities is significantly higher than in neighbouring regions could as a whole be more relevant, as it offers more opportunities for a user to perform a related activity. Taking into account the neighbourhood of a geographic entity provides further information about the spatial context of that location and could for instance identify or link to associate places or regions, as well as co-located entities. In this perspective, first- and second-order effects in the distribution of the entities in the geographic space (O'Sullivan and Unwin, 2003) can be analysed and geographic data mining techniques (Miller and Han, 2009) can be exploited, in order to understand complex geographic phenomena that involve not only a single geographic entity, but also its surroundings. This information plays an important role in spatially- and context-aware concepts of relevance such as GR (De Sabbata and Reichenbacher, 2012).

Finally, a semantic model of space addresses the issue of the specific meaning space may have (e.g., places, regions, functions, and qualities of places). On the one hand, most users of a mobile service think of location in space in terms of places that have a specific meaning for them (Gibson, 1986). This meaning may be constructed through experiences. On the other hand, places may also offer specific functions or afford certain activities (Jordan et al., 1998; Alazzawi et al., 2010; Goodchild, 2011). Hence, the mobile activity a user wants to perform constrains the relevance of places or regions, based on affordances and meaning assigned to the latter. For instance, in order to perform the activity of dining, one needs a place that offers the possibility to eat, therefore restaurants are more relevant than garment shops.

3.3 Conceptual model

In this section I present a conceptual model, in which I summarise my approach with respect to the first research question (RQ1), as formulated in Section 1.2. The inception is the concept of GR as defined and described above. In order to analyse this complex relationship, it is necessary to first establish a model for the information involved in the GR assessment, and related to the user, the geographic entities, and the surrounding environment. This allows to investigate the relationships between pieces of information describing the user and pieces of information describing geographic entities, in the context of information available about the surrounding environment.

To achieve this purpose, the components listed in the framework described above (see Section 3.1) can be used. In fact, each component defines an aspect of the first three dimensions of relevance (i.e., information need, information resource, and world representation). Thus, the information required to model the user context, geographic entities, and the surrounding environment is then identified by analysing the corresponding relevance dimension by means of the defined components.

The conceptual model of GR (see Figure 3.2) includes two main parts, which entail the information needed to model the user, and the geographic entities. Further elements are included in these two parts, which describe the system's representation of the world that surround both user and geographic entity respectively. The user is described as a combination of the components activity, preferences, social context, and mobility.

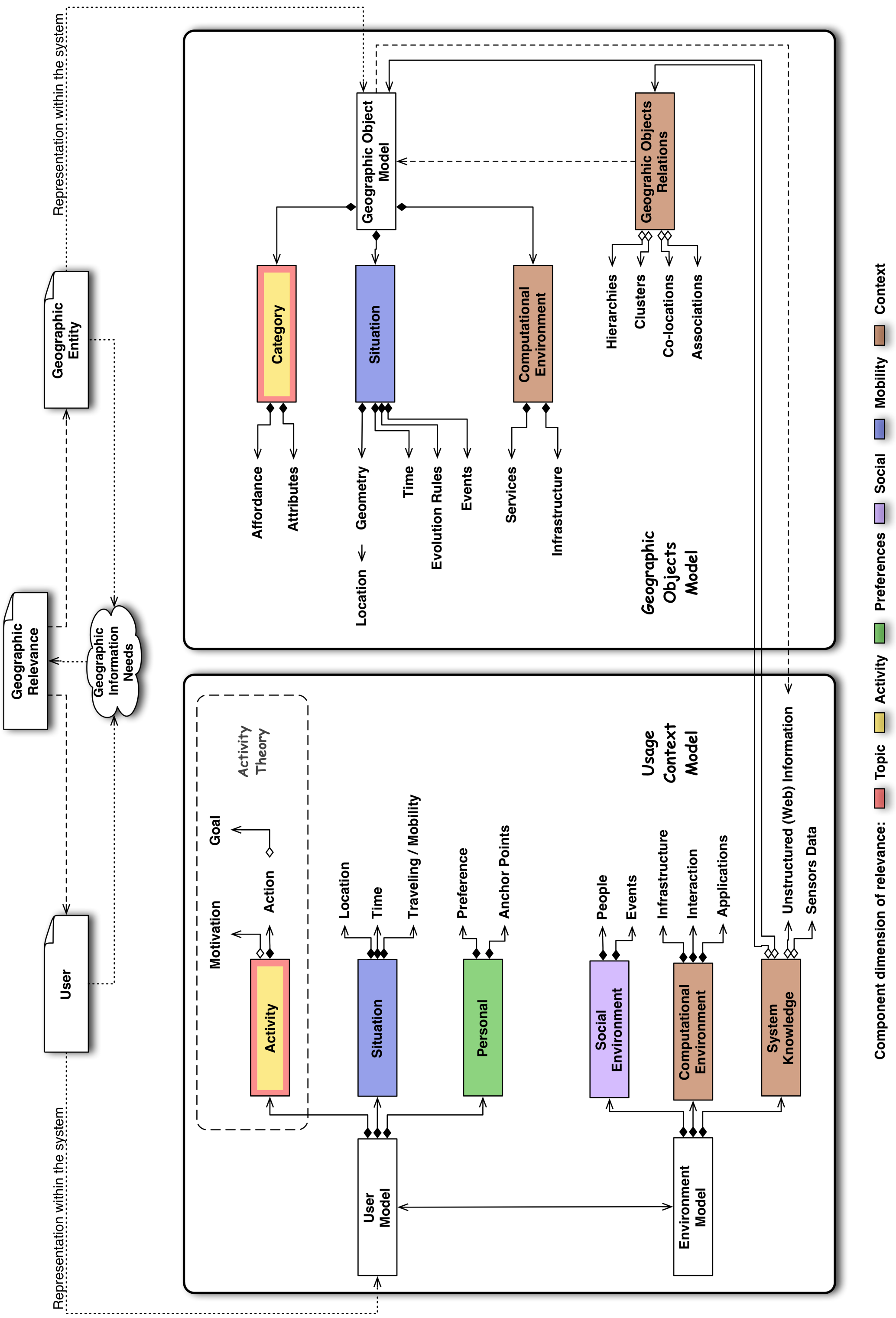


Figure 3.2: Conceptual model of geographic relevance.

The geographic entity is mainly described as a combination of the components activity and mobility. The system's representation of the world completes the model adding contextual elements related to the informational and geographic surroundings of both user and entity.

The central part of the model (in the top level in Figure 3.2) reflects the fundamental idea of GR as a relation between a user in a given location and context, and geographic information entities as representations of real world entities. The first part (on the left in Figure 3.2) is the user model. It includes a description of the user's current situation (location and time), and mobility information.

Moreover, her preferences can be modelled, along with her anchor-points (Couclelis et al., 1987), which can be understood as salient places in her understanding of the geographic space. That is, places that she visits frequently, or stays at for long periods (e.g., her home or work place), or popular landmarks. Furthermore, the user model includes her activity and actions, as well as her motivations and goals. These also convey information about the topic in which she is interested.

The user model is associated with a model of the current environment surrounding the user, including her social environment (e.g., when travelling in a group), the computational environment of her device, and the system's knowledge of reality. The latter represents the system's knowledge about the world (Zipf and Jöst, 2006) and encompasses information about built-in device sensors (e.g., temperature sensors), remotely accessible information (e.g., Internet services), and surrounding geographical entities. It is also linked to a set of relations among these entities capturing first- and second-order effects in the distribution of the entities in the geographic space (see e.g. O'Sullivan and Unwin, 2003), which is included in the second part of the model as described in the following.

The geographic entities, together with the relations between them, are described by the second part of the model (on the right in Figure 3.2). On the one hand, a geographic entity has to be identifiable by a numerical ID or unique name. On the other hand, the category of the entity (e.g., restaurant) has to be modelled, including attributes information (e.g., restaurant rating), as well as the computational environment (i.e., infrastructural and software services). Taking into account a semantic model of space, affordances or functions (e.g., offered services) are also included, since this information is fundamental in understanding the components of GR related to a user's activity and topic of interest. Each entity is also described by its geometry, defining its location, and it may possibly refer to a place.

Moreover, since a semantic model of space embodies a structural model, the geographic relations among entities are also included (i.e., first- and second-order effects). Geographic relations taken into account include (but are not limited to) general topological relations, spatial hierarchies, clusters, co-locations, connectivity, and spatio-temporal association rules. This information is necessary to understand the geographic context component of GR. Finally, the model includes time-related information. The local time of the entity (e.g., current time, day time, or season) is modelled, along with related events (e.g., plays of a theatre). Evolution rules describe the manner in which entities change over time (e.g., opening hours of shops).

3.4 Summary

In this chapter, I presented the concept of GR, which is the main focus of this dissertation. I also proposed a new framework, which can be used to describe conceptualisations of relevance in four dimensions and a set of components of relevance. This framework was used to compare GR with other concepts of relevance proposed in the fields of IR and mobile information services. The components of relevance defined in this framework were used to analyse GR, and identify the information needed in order to model the user context, geographic entities, and the surrounding environment. The result is summarised in the conceptual model of GR, presented in Figure 3.2.

GR strongly emphasises the geographic facet of relevance. Thus, the assessment of GR is closely tied to the selected conceptualisation of space, which can have a strong influence on the assessment results. A topological model of space can convey a more realistic understanding of the user's mobility, with respect to a geometrical model of space. The structural and semantic models can enable an even deeper understanding of the context in which the information seeking process happens.

This chapter offers an answer to the first research question (RQ1, see Section 1.2), clarifying the elements involved in the GR relationship. The next chapter will focus on which criteria of relevance define the GR relationship, and therefore have to be taken into account in the assessment of GR, proposing an answer to the second research question (RQ2, see Section 1.2).

Chapter 4

Criteria

In the previous chapter, I defined the concept of GR and presented a conceptual model of GR, describing the components involved in the assessment of the GR of a geographic entity for a mobile user in a given context. This raises the challenge to identify an appropriate set of criteria that is capable of operationalising those components.

In this chapter, I investigate whether the criteria developed in IR and related fields (see Chapter 2) have to be taken into account for assessing GR (or if some of them do not apply) and whether those criteria are sufficient (or if there are other criteria which have not been investigated yet[1]). The guiding question is whether the concept of GR is equivalent to the concept of relevance employed in IR, or not.

In Section 3.1, GR was defined as a derivation from the concept of situational relevance proposed in IR. It was also illustrated how GR is quite similar to the concepts of relevance used in GIR and mobile information services. This suggests that at least some criteria proposed in the IR, GIR, and mobile information services literature (see Section 2.1 and 2.3) should be applicable to GR. Many conceptual differences have already been highlighted in Chapter 3, in particular the focus of GR on the geographic component of relevance, and the use of a semantic model of space. This implies that specific criteria might be important for assessing GR, especially to achieve an adequate understanding of the user's mobility and the geographic environment. Although these are important conceptual facets of GR, it can not be excluded that these factors might have little actual influence on the users' perception of relevance compared to the more classic criteria of relevance (e.g., topicality, appropriateness).

This chapter starts with a description of criteria, that could be applicable to GR. Section 4.1 describes how criteria developed in IR and mobile information services can be adapted to GR. Thereafter, I present new criteria specifically proposed in the scope of GR, which should also be considered in assessing GR. The three following sections present two empirical studies (i.e., Experiment I and II), which I conducted to investigate the importance of the listed criteria, and a discussion of the results. These sections are largely based on De Sabbata and Reichenbacher (2012). Finally, Section 4.4 summarises the findings of the experiments and sets the stage for the next chapters.

[1]This chapter is largely based on De Sabbata and Reichenbacher (2012).

Table 4.1: Criteria of GR.

Properties	Geography	Information	Presentation
topicality	spatial proximity	specificity	accessibility
appropriateness	temporal proximity	availability	clarity
coverage	spatio-temporal proximity	accuracy	tangibility
novelty	directionality	currency	dynamism
popularity	visibility	reliability	presentation quality
	anchor-point proximity	verification	
	hierarchy	affectiveness	
	cluster	curiosity	
	co-location	familiarity	
	association rules	variety	

4.1 Definitions

A first list of criteria of GR has been proposed in De Sabbata (2010), and then refined in De Sabbata and Reichenbacher (2012). Table 4.1 presents the refined list of criteria of GR taken into account in this dissertation. The criteria are grouped in four classes as suggested in De Sabbata (2010).

Barry and Schamber (1998) advocate for '*the existence of a finite range of criteria that are applied across types of users, information problem situations, and information sources*', whereas the few differences '*appear to be due to the differences in situational contexts and research task requirements*' (Barry and Schamber, 1998, p. 234). This is consistent with the hypothesis suggested above, that most of the criteria used in IR and related fields would apply to GR, while other distinctive criteria may be needed.

The class *properties* includes criteria which refer to the properties of an entity. These criteria are drawn from criteria developed in IR (see Section 2.1) and their descriptions are reported in Table 4.2. These criteria relate to the topic, activity, and preferences component of the framework presented in Section 3.1. The implementation of these criteria mainly deals with the activity and personal elements of the user component, the category element of the entity component, and the computational environment elements of both user and entity components of the conceptual model of GR (see Section 3.3 and Figure 3.2). In particular, the criteria coverage and appropriateness are strongly related to the affordance of a geographic entity. A semantic model of space is needed in order to implement these criteria.

All criteria within the class *properties*, except *popularity*, which is similar to the concept of page popularity in the Web, were presented in (De Sabbata, 2010). Page popularity has been largely exploited since the introduction of PageRank (Page et al., 1999), using the link citation between pages as a proxy. The understanding of this criterion is changing, due to the emergence of social networks. For instance, the criterion popularity is widely used in recommender systems, which give users recommendations about objects that are popular among 'similar' users. The same approach can be taken into account in assessing GR. Therefore this criterion involves the social environment element of the conceptual model of GR.

The class *geography* includes criteria that refer to the geographical essence of an entity. The first half of Table 4.3 (above the horizontal line) describes the criteria related to the user's mobility. These criteria refer to the mobility component of the framework presented in Section 3.1. These criteria involve the location and time elements of both the user and the entity components of the conceptual model of GR. All these criteria were discussed in Section 2.3, with the exception of anchor-point proximity, which was proposed in the scope of GR in De Sabbata and Reichenbacher (2012). Anchor-point proximity is not strictly related to a user's current mobility, but more to her habits or to a wider understanding of a place. In fact, the user's home or her workplace can be considered anchor-points, along with popular landmarks and highly frequented places (Couclelis et al., 1987). This criterion is therefore linked to the social component of GR. For instance, Venetis et al. (2011) suggested to rank spatial Web object based on the frequency with which places get mentioned in queries for driving direction on Google Maps. However, anchor-point proximity has an individual or personal meaning, whereas the direction-based approach targets the popular landmarks and highly frequented places.

The second half of Table 4.3 outlines criteria that have been specifically proposed within the scope of GR (De Sabbata, 2010), and which refer to fundamental concepts in geography. These criteria are concerned with the geographic context in which the entities are placed, and involve the context component of the framework presented in Chapter 3. The underlying idea is that geographic entities do not exist as simple objects in an empty Euclidean space, independent from each other. Geographic entities are more complex, they exist within a specific geographic environment, and they are part of geographic phenomena characterising that environment, such as spatial clusters or co-location patterns. These criteria pertain to a structural model of space and aim to capture such geographic phenomena and (see Section 3.3) by means of the data mining techniques described in Section 2.4, and take them into account when assessing relevance. These criteria deal with the geographic entities relationships element of the conceptual model of GR, but also with the situation element of the user component in the conceptual model of GR (see Figure 3.2).

The class *information* includes criteria that refer to the quality and availability of the information representing geographic entities within the system. The class *presentation* includes criteria which refer to how well the available information is presented to the user. The criteria included in both of these classes derive from criteria proposed in the scope of document-based IR, and were discussed in Section 2.1. In Table 4.4 and 4.5, I propose an adapted version of each criterion, which can be considered as possible criteria of GR.

Therefore, the list of possible criteria of GR includes most of the criteria developed in IR and mobile information services. These are enumerated side by side with five new criteria, that I suggested in the scope of GR (i.e., the criterion anchor-point proximity, and the four criteria at the bottom of Table 4.3). The question is whether these five new

Table 4.2: Criteria definitions: properties.

Criterion	Definition
Topicality	The extent to which the category of the entity matches the user's needs in accomplishing an activity.
Appropriateness	The extent to which the affordance of the entity is focused on the user needs.
Coverage	The extent to which the user needs are satisfied by the affordance of the entity.
Novelty	The extent to which the entity or related information are unknown or novel to the user.
Popularity	The extent to which the entity is popular or highly regarded by a community, or among a group of users 'socially' related to the user's.

Table 4.3: Criteria definitions: geography.

Criterion	Definition
Spatial proximity	The extent to which the entity is spatially close to the user's location.
Temporal proximity	The extent to which an entity (or an associated event) is temporally close to the user.
Spatio-temporal proximity	The extent to which the entity (or a related event) is spatio-temporally close to the user – it may be past, current, or upcoming at the time the user will be at the location of the entity – and how long this status will last from the moment the user will have arrived at that location.
Directionality	The extent to which an entity is in the same direction the user is heading, or the amount of detour needed to include the location of the entity in the path planned by the user.
Visibility	Whether the entity can be seen from the user's location or not.
Anchor-point proximity	The extent to which the entity is spatially close to a place that the user accounts as an anchor-point – e.g., a place that the user visit frequently or where the user spend a lot of time.
Hierarchy	Degree of separation between the po- sition of the user and the location of the geographic entity within a predefined spatial hierarchy.
Cluster	Degree of membership of an entity to a spatial cluster of related or unrelated entities – the size of the cluster can also be taken into account as a factor of relevance.
Co-location	The extent to which an entity satisfies a co-location pattern, that has been identified as common and meaningful for that category of entities within a related collection of geographic entities, and apropos the user's current needs.
Association rules	The extent to which an entity satisfies an association rule, that has been identified as common within a related collection of geographic entities – these rules can involve spatial, temporal, and/or other types of attribute.

Table 4.4: Criteria definitions: information.

Criterion	Definition
Specificity	The extent to which the information about an entity has sufficient detail or depth.
Availability	The extent to which information or sources of information about the entity are available to the user through the information system.
Accuracy	The extent to which the information about an entity is accurate, correct or valid.
Currency	The extent to which the information about the entity is current, recent, timely, up-to-date.
Reliability	The extent to which general standards of quality or specific quality standards can be assumed, based on the source providing the information; source is reputable, trusted, expert.
Verification	The extent to which information about an entity is consistent with or supported by other information on the same subject.
Affectiveness	The extent to which the user exhibits an affective or emotional response to information or source of information; that is the extent to which the information or the sources of information provide the user with pleasure, enjoyment or entertainment.
Curiosity	The extent to which access to information is dependent on personal curiosity.
Familiarity	The the extent to which the user is familiar with the source of information.
Variety	The extent to which the source provides a sufficient variety of information.

Table 4.5: Criteria definitions: presentation.

Criterion	Definition
Accessibility	The extent to which some effort or cost is required to obtain information (not to be mistaken with the concept of spatio-temporal accessibility in GIScience, as explained later on).
Clarity	The extent to which the information about the entity is presented in a clear and well-organized manner.
Tangibility	The extent to which the information presented to the user relates to real, tangible issues; definite, proven information is provided; hard data or actual numbers are provided.
Dynamism	The can be defined as the extent to which presentation of information is dynamic, active or live (e.g., presentation manipulation, zooming).
Presentation quality	The extent to which a source presents information in a certain format or style, or offers output in a way that is helpful, desirable, or preferable (choice of format, entertainment value).

criteria are necessary, and whether they are enough to assess GR, or not. To answer these questions, two experiments are proposed. The null hypothesis is that the information overload problem can be solved by applying relevance criteria developed in the field of IR to the information pieces identified as key components for GR. This would be rejected if the results of the two experiments indicate that the five new criteria are important for the assessment of GR. This would also confirm the existence of those few differences mentioned by Barry and Schamber (1998), and advocate for GR as a separate and new concept of relevance in GIScience.

4.2 Experiment I

Experiment I was designed as a Web-based questionnaire, and it aims to gain a first insight into the importance of a subset of the criteria presented in the previous sections. The main interest is to test the importance of four newly proposed criteria related to the geographic context described in the second half of Table 4.3: hierarchy, cluster, co-location and anchor-point proximity.

The two following sections briefly describe the methods and discuss the results of Experiment I. Further information is reported in Appendix A. This experiment is discussed in detail in De Sabbata and Reichenbacher (2012).

4.2.1 Method

Participants. A total of 132 participants took part in this experiment. A first group of 53 participants included researchers gathered through mailing-lists dedicated to topics in computer science and geography, but also non-academics individuals. This first group participated in a web survey developed using the online service SurveyMonkey[2]. The second (39 participants) and the third (40 participants) groups were gathered through the online service Amazon Mechanical Turk[3]. These participants were assumed to fall into Amazon Mechanical Turk's demographics (Ross et al., 2010) of computer literate people with no particular expertise in geography.

Scope. The overall idea of this first study was to ask participants about their opinions on the usefulness of the criteria described in Section 4.1. Not all criteria listed in Table 4.1 were taken into account, since the aim was to focus specifically on the geography-related criteria.

Materials. I developed three similar web-based on-line questionnaires: one was developed using the online service SurveyMonkey, and two were developed using the online service Amazon Mechanical Turk. Further details about the structure and content of the three questionnaires are reported in Appendix A.

Procedure. The first page of each questionnaire stated the objective of the project and the purposes of the study. Then, participants were asked whether they agree or

[2] `http://www.surveymonkey.com`, last accessed November 2012.
[3] `https://www.mturk.com/mturk/welcome`, last accessed November 2012.

disagree (on a 7-point Likert scale) with 15 statements, each one representing one of the criteria taken into account. The statements used in this experiment are derived from the definitions presented in Section 4.1, and are reported in Appendix A.

4.2.2 Results and discussion

The results of Experiment I clearly indicate that the participants agree with the usefulness of the geographic criteria. In particular, a promising level of agreement on the usefulness of the four recently proposed criteria (i.e., hierarchy, cluster, co-location and anchor-point proximity) was observed. These were rated as important factors in the judgement of the geographic relevance of an entity (see Table 4.6).

The highest rated criteria are spatio-temporal proximity and coverage with mode equal to the highest score (see Table 4.6). The participants also 'agree' on the importance of the criteria currency, accuracy, and anchor-point proximity. The majority of participants at least 'somewhat agree' (with 'agree' as the most common opinion) on the importance of a group including four geographic criteria (i.e., co-location, hierarchy, directionality, and cluster), and the criteria availability and appropriateness. Finally, the criteria presentation quality and visibility were rated lower in importance, and the criteria dynamism and novelty got the lowest scores. The participants seem to just 'somewhat agree' with the former, and seem to be 'neutral' with respect to the latter.

These results provide a first insight into the applicability of the single criteria of GR and a first confirmation of the importance of the criteria related to the geographic context described in the second half of Table 4.3. This, in turn, suggests that the geographic facet of this retrieval problem appears to be significant, and a clear indicator of a difference between GR and the concept of relevance employed in classic document-based IR.

A substantial difference between GR and classic document-based IR is also reflected by the rates given to the criteria presentation quality and novelty. The first was the most mentioned criterion in Schamber's study of criteria of relevance (Schamber, 1991), and the second was the third rated in the output list of Barry (1994). The results of this survey indicate that, when a user has to judge a geographic entity rather than a document, these criteria can be accounted as secondary, maybe even optional. The same applies to the criterion dynamism, that can be also found in Schamber's list of criteria of relevance (Schamber, 1991). Moreover, this difference is confirmed by the agreement about the usefulness of the criteria anchor-point proximity, co-location, hierarchy, and cluster. These four new geographic criteria are a distinguishing feature of the retrieval of geographic information, and they seem to play an important role in GR.

Finally, a statistically significant difference ($p < .01$) was found for the median of the rates collected with the first questionnaire for the five criteria availability, accuracy, dynamism, presentation quality and visibility, with respect to the median of the rates collected with the second and the third questionnaire. That is, a statistically significant difference was found between the responses given by the participants to the first Web-based questionnaire (mostly researchers and students in GIScience and Information

Table 4.6: Mode and median values of the responses collected in Experiment I.

Criterion	Mode	Median
Spatio-temporal proximity Coverage	Strongly agree	Agree
Currency Accuracy Anchor-point proximity	Agree	Agree
Availability Co-location Hierarchy Directionality Cluster Appropriateness	Agree	Somewhat agree
Presentation quality Visibility	Somewhat agree	Somewhat agree
Dynamism Novelty	Neutral	Neutral

Retrieval) and the responses collected using Amazon Mechanical Turk. No statistical difference has been found between the data collected with the second and the third questionnaire. A detailed analysis of these differences is presented in (De Sabbata et al., 2012).

The experiment described above has two main limitations. First, there may be a difference between the answers given by a participant when asked about a criterion and the actual usage of the criterion. In fact, the role of a criterion may not be clear until one has to use it in a practical situation. Second, different participants might have very different situations in mind when answering the same question, which can influence their answers. In the next section, a second experiment is presented where each participant is faced with an explicit mobile usage context, and given geographic information needs.

4.3 Experiment II

Experiment II was designed to establish whether the geographic context of entities has a significant impact on their GR. The underlying research question was whether similar geographic entities at similar distances from the user's position would get different relevance judgements if placed in different geographic context. In particular, Experiment II tested three newly proposed criteria related to the geographic context, as described in the second half of Table 4.3: co-location, hierarchy, and cluster.

The two following sections briefly describe the methods and discuss the results of Experiment II. Further information is reported in Appendix B. This experiment is discussed in detail in De Sabbata and Reichenbacher (2012).

4.3.1 Method

Participants. A total of 110 participants took part in this experiment. The participants were gathered by sending e-mails to different mailing-lists, Google Groups[4], and Yahoo Groups[5], related to the fields of IR, GIScience, and cartography. I assumed that participants gathered by those means would have at least some familiarity with web search engines and digital maps. Participants were randomly assigned to one of four sub-scenarios.

Scope. Two different scenarios were developed for Experiment II, each one composed of two sub-scenarios. The second sub-scenario of each scenario included additional information not included in the first sub-scenario. This supplementary information was intended to allow the participant to apply one or more additional criteria with respect to the first sub-scenario. The objective was to compare the usage of the criteria between the different groups of participants (each one responding to a different sub-scenario).

The four sub-scenarios were presented to different groups of participants. The participants were asked to simulate (i.e., "act as", "play the role of") a hypothetical GR assessment system, taking into account all available information and the criteria they consider to be important, in order to judge the relevance of the individual geographic entities. The aim was not to test an actual application such as a geographic recommendation system or LBS.

Materials. A detailed description of the material used in this experiment is presented in (De Sabbata and Reichenbacher, 2012) and reported in Appendix B, including the text and the maps presented to the participants.

In the first scenario, the participant was told that she was looking for a hotel for the night, in a city where she has never been. In the first sub-scenario (S1A, see Appendix B), the hypothesis was that the participant would take into account the presence of restaurants, museums and tourist attractions, where the hotels near those POI are more relevant than the others. In the second sub-scenario (S1B), some of the representations of hotels on the map were accompanied by some further information on the price of the room or a hotel picture. The hypothesis was that the participant would consider those hotels more relevant than the others – still holding the previous hypothesis.

In the first scenario, the participant was told that she was looking for a restaurant for the night, in a city she has never been before. In the first sub-scenario (S2A), the hypothesis was that the participant would take into account the opening hours of the restaurants – excluding those restaurants currently closed or about to close – and the visible clusters of restaurants, where the restaurants that are part of a cluster would be more relevant than the others. In the second sub-scenario (S2B), the participant was told that she has to go back to the hotel after lunch. The hypothesis was that the participant would consider the restaurants along the route to her hotel to be more relevant.

[4] `http://groups.google.com`
[5] `http://groups.yahoo.com`

Procedure. On the first page of the questionnaire, the purpose of the study was stated. On the second page, the scenario and a related map was presented (see Figures B.1 and B.2). The participants were asked to rate the relevance on a scale from 1 to 7 (i.e, 1 = "not relevant at all", 4 = "somewhat relevant", 7 = "extremely relevant") of a set of objects displayed on the map and to give a brief mandatory description that explained their ratings. On the third page, the participants were asked whether they used the hypothesised criteria. As with Experiment I, sentences derived from the definitions presented in Section 4.1 were used to the represent the hypothesised criteria. The statements used on the third page of Experiment II are reported in Appendix B. An optional comment box was provided on the third page. The four questionnaires for the four sub-scenarios were developed using the online service OnlineUmfragen[6]. The following section describes in detail the composition and purpose of the different scenarios and sub-scenarios.

4.3.2 Results and discussion

The key aspect of the obtained result is that similar geographic entities at similar distance from a user's position do receive different relevance judgements, if placed in different geographic settings. The responses collected in this experiment confirm the insights gained from the first experiment (see Section 4.2), and suggest a rejection of the hypothesis of equivalence between GR and the concept of relevance employed in IR. The results also confirm the importance of the three geographic criteria tested, and reassert the uncertainty about other criteria. A detailed description of the results is presented in (De Sabbata and Reichenbacher, 2012).

The importance of the criteria co-location and cluster clearly emerges from the results and is supported by comments obtained from the first and second scenario respectively. The closeness of a hotel to points of interest, such as restaurants and museums, seems to be a good criterion to identify highly relevant hotels. In all scenarios, the participants favoured those entities which are located in the city centre at the expense of entities located in residential areas. This suggests that the criterion hierarchy is somewhat important, but it seems to be superfluous in the second scenario, where most participants took into account this criterion, but decided not to use it – the use of the other criteria seems to be enough to make a decision about the relevance of an entity.

The outstanding importance of the criterion spatio-temporal proximity is evident in the results obtained in the the second scenario, where the restaurants which are closed or about to close have been clearly rated as not-relevant. Directionality also plays an important role, leading to a prominent difference between the two sub-scenarios of the second scenario. The ratings of those restaurants which are nearby the hotel change from slightly relevant to top-ranked, when the participant is told about her further destination.

In the first scenario, the importance of the criterion availability is also very clear.

[6] http://www.onlineumfragen.com

Most of the participants rated the hotels with further information as more relevant than the others. Instead, the role of the criteria accuracy and presentation quality is rather unclear. Roughly a quarter of the participants used the first criterion, one third of the participants stated to have used the second criterion, and one third of the participants stated that they would use them, but did not think about it. Nonetheless, it is also difficult to unquestionably distinguish the influences of the these two criteria from the influence of the criterion availability, and these three criteria are not fully independent from one another. Still, the effect of accuracy and presentation quality criteria on the overall relevance of the entity seems to be rather narrow.

Combining these results with the results from Experiment I, it becomes clear that there are some criteria which determine whether an entity is relevant or not. In the scope of GR, these "fundamental" criteria are topicality or spatio-temporal proximity, which can be used to filter out options that do not fit at all the user's needs in terms of user's interest or mobility limitations. A second set of "primary" criteria is composed by those that define how much relevant a feasible option is. This includes those criteria that implied a significant difference in the rates, such as availability, directionality, co-location, cluster, and hierarchy. A last, third set of 'secondary' criteria is composed by those that can help to distinguish between two similar entities, which are however non-compensatory criteria – that is, they do not have a significant impact on the relevance of an entity. The criteria presentation quality and accuracy fall into this set. The importance of these last criteria may be related to the context of the search or personal preferences.

Finally, the analysis of the explanations given by the participants to justify their responses confirm that the class *geography* (see Tables 4.1 and 4.3) is not homogeneous. That class is composed of two distinct sub-groups of criteria, one related to the user's mobility and the other related to the environment surrounding the geographic entities. These are criteria that are commonly mentioned together in the explanations, sometimes they are combined to form a more general concept.

4.4 Summary

The results obtained from Experiment I and II (see also De Sabbata and Reichenbacher, 2012) showed that GR and the concept of relevance commonly employed in IR are different. In fact, the mobile and geographic context components of relevance concern the first concept, but are not of concern to the second. It is clear that the criteria co-location, cluster, and hierarchy (see Table 4.3) are among the primary criteria of GR. Hence, assuming that the criteria topicality and spatio-temporal proximity are satisfied, the geographic context of the entities has to be taken into account in order to assess GR. The findings also suggest directionality and anchor-point proximity to be a valuable contributor to understand the user's mobility. This answers the second research question (RQ2) in Section 1.2.

Therefore, the classic concept of relevance can not be applied in a naïve way to systems dealing with geographic entities instead of documents. The conceptual model presented in Chapter 3 has to be complemented with the criteria identified as important in Experiment I and II, in order to form the conceptual base of a new computational model for the GR assessment. In Chapters 5 and 6, I suggest a framework to compute numerical scores for a selected set of criteria of GR, based on the information encapsulated in elements described in the conceptual model of GR. The computational framework is then evaluated in Chapter 7.

Chapter 5

Relevance assessment

In this chapter, I propose a computational model for the assessment of GR, to implement five of the criteria empirically validated in the previous Chapter 4, based on the conceptual model developed in Chapter 3. The proposed model delineates a procedure to quantitatively estimate GR, aggregating different scores calculated for a set of five selected criteria.

Section 5.1 discusses the selection of the criteria to be included in the computational model, their links to the conceptual model (see Chapter 3), and their contribution to the quantitative assessment of GR. Sections 5.2 and 5.3 provide an exploratory answer to the third research questions (RQ3, see Section 1.2), issuing formal definitions of the numerical scores associated with the selected criteria. Finally, Section 5.5 provides an exploratory answer to the fourth research question (RQ4, see Section 1.2), discussing various methods to combine the numerical scores defined in the previous sections in an aggregate value of GR.

The aim of this chapter is to present the GR assessment model, and provide the formal definitions needed for the software implementation. The latter is described in further detail in Chapter 6. The evaluation of the proposed computational model is then presented in the Chapter 7.

5.1 Assessment model

I formalise the quantitative GR assessment model as a combination of five criteria of GR, based on the empirical results presented in Chapter 4. Prior study supports that the criteria topicality and spatio-temporal proximity are the most fundamental criteria of GR, thus these criteria have to be part of the computational model. Experiment II (see Section 4.3) additionally supports the importance of the criterion directionality, thus this criterion should also be included. At the same time, the criterion directionality is not completely independent from the criterion spatio-temporal proximity, as both criteria take into account the user's destination. This issue is further discussed in Sections 5.3.2 and 5.3.3, when the definitions of these two criteria are given.

In addition to these first three criteria, I include the criteria cluster and co-location

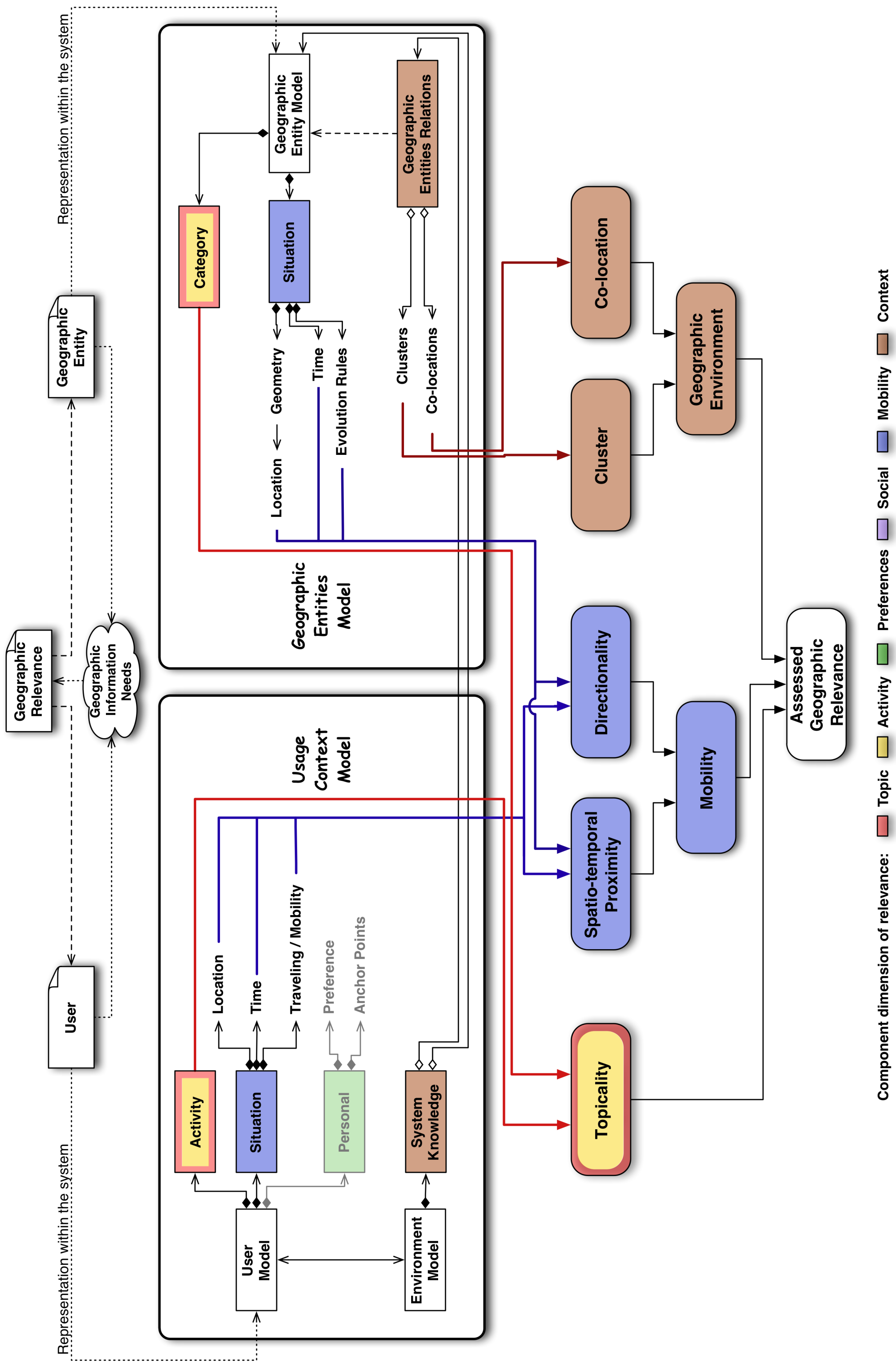


Figure 5.1: Computational framework for the assessment of GR.

in this assessment model of GR, at the expense of more traditional IR criteria. There are three compelling reasons for this. First, the results presented in Chapter 4 show how cluster and co-location are among the primary criteria of GR. Second, I consider it important to deepen the understanding of two recently proposed criteria (i.e., cluster and co-location), which have never been tested before (see De Sabbata and Reichenbacher, 2012). Third, these criteria convey information about the geographic environment of the entities, which is a distinguishing component of GR, and is not part of the concept of relevance as it is commonly understood in IR, GIR, and MIR. In fact, the criteria cluster and co-location are one of the main discerning factors of GR (see Chapter 4). A step in this direction has been undertaken by Ostuni et al. (2012), who implement the five criteria as proposed in De Sabbata and Reichenbacher (2012) (i.e., hierarchy, cluster, co-location, association-rule, and anchor-point proximity) as binary filters within the "Cinemappy" mobile application for movie recommendations. In this dissertation, I present, implement, and empirically evaluate a multi-facet, non-binary definition of the criteria cluster and co-location (see Sections 5.3.4 and 5.3.5, respectively). Future research will focus on developing an analogous definition for the numerical estimation of the criteria hierarchy, association-rule, and anchor-point proximity.

The computational model is based on three assumptions. First, for each criterion, it is possible to compute a 'distance value', which estimates the distance or difference between the user's information need and an entity under relevance assessment, with respect to the criterion under consideration. A semantic distance is taken into account for the criterion topicality. Spatial and temporal distances are taken into account for the criteria spatio-temporal proximity and directionality. Spatial distances and numerical differences (e.g., difference in the cardinality of two sets of objects) are taken into account for the criteria cluster and co-location. Second, based on the 'distance value' computed for a given criterion, it is possible to compute a 'score' (i.e., a numerical value normalised in the interval $[0, 1]$), which estimates the strength of GR, with respect to the criterion under consideration. This understanding of GR is inspired by the first law of geography (Tobler, 1970), which is extended from the conventional geographic space to the non-geographic space of each criterion. For instance, concepts which are near in the semantic space (i.e., concepts whose semantic distance is small) are assumed to be more related than concepts that are distant in the semantic space (i.e., concepts whose semantic distance is large). Hence, the lower the 'distance value' computed for a criterion, the higher the 'score'. Third, a numerical estimation of GR can be computed combining the 'scores' described above. These assumptions are taken into account within the scope of assessing GR as defined earlier in this dissertation (see Section 3.1).

Therefore, GR is assessed through a numerical estimation and combination of the criteria topicality, spatio-temporal proximity, directionality, cluster, and co-location. In turn, the numerical estimation of each criterion is grounded on the conceptual model presented in Section 3.3 (see Figure 3.2), as illustrated in Figure 5.1. The top half of the illustration reports a reduced version of the conceptual model presented in Figure 3.2. The bottom part in Figure 3.2 depicts how GR is derived from the elements of the conceptual model by means of the selected criteria.

Topicality takes into account a user's activity, and the category an entity belongs to, in order to achieve an estimated score for the topic and activity components of GR (see Section 3.1). Spatio-temporal proximity and directionality account for the situational elements in the user model and geographic entities model, in order to derive the respective estimated scores. These two scores are combined into an estimated score for the mobility component of GR. Cluster and co-location account for the relationships among geographic entities, which are the part of the environment model included in the geographic entities model, in order to derive the respective numerical scores. These last are combined in an estimated score related to the geographic environment of entities, that is part of the context component of GR. Finally, the scores computed for each component are aggregated into a final GR score, according to their relative importance, as discussed in Chapter 4.

5.2 Base definitions

The definitions presented in the following sections will refer to the basic elements listed below, which will be assumed as available, independent of the underlying implementation. This formalisation is introduced in order to avoid long and ambiguous descriptions of the functions used, although a fully grounded mathematical definition of the problem is out of the scope of this thesis.

$$
\begin{aligned}
q \in Q &: \text{ user query} \\
G = \{g_1, g_2, \dots\} &: \text{ geographic entities} \\
cat(g) &: \text{ category which } g \text{ belongs to} \\
dist(g_1, g_2) &: \text{ spatial distance between entities } g_1 \text{ and } g_2 \\
\Phi = \{\phi_1, \phi_2, \dots\} &: \text{ clusters of geographic entities} \\
\phi(g) &: \text{ cluster which } g \text{ belongs to} \\
\Phi^{cat(g)} = \{\phi_i, \phi_{i+1}, \dots\} &: \text{ clusters of geographic entities belonging to} \\
&\quad \text{ the same category as a given entity } g \\
\Psi = \{\psi_1, \psi_2, \dots\} &: \text{ co-location rules regarding the geographic entities} \\
&\quad \text{where } \psi^p \text{ is the ``premise'' of a rule } \psi \\
&\quad \text{and } \psi^c \text{ is the ``conclusion'' of a rule } \psi \\
\Psi(g) = \{\psi_j, \psi_{j+1}, \dots\} &: \text{ co-location rules having as ``premise''} \\
&\quad \text{ the same category to which } g \text{ belongs to} \\
t_{Clust} &: \text{ threshold used for mining the clusters} \\
t_{Coloc} &: \text{ threshold used for mining the co-location rules}
\end{aligned}
$$

The user query element includes the user's current location, her destination, the activity for which she needs information, including the minimum time needed to perform this activity. The geographic entity elements contain all available information about the entities, including their category, location, and temporal availability. The remaining elements refer to the information obtained mining the dataset for spatial clusters and co-location rules (see Sections 2.4 and 4.1).

As from the assumptions advanced in Section 5.1, for each of the selected criteria, it is possible to compute a "distance function" δ for each criterion, such as:

$$\delta_{Criterion} : Q \times G \rightarrow \mathbb{R}_0^+ \tag{5.1}$$

which takes a user query and a geographic entity as input. The output value grows as the relevance of the geographic entity for the user context declines, in the scope of the considered criterion. Furthermore, for each of the selected criteria, it is possible to compute a "normalised score" such as:

$$\overline{s}_{Criterion} = f \circ \delta : Q \times G \rightarrow [0 \ldots 1] \tag{5.2}$$

which is a function of the "distance function" defined above. Such a function takes a user query and a geographic entity as input, and returns a value between 0 and 1. The value 1 is assigned to the most relevant geographic entity for the user query in the scope of the considered criterion, and the value 0 is assigned to geographic entities which are completely irrelevant for the user query in the scope of the considered criterion.

5.3 Criteria scores

In the following paragraphs, the definitions of the distance functions δ and score functions $\overline{s}$ are issued for each criterion, along with a series of auxiliary functions d. This section provides an exploratory answer to the third research question (RQ3) in Section 1.2.

5.3.1 Topicality

The aim of this work is to achieve a numerical estimation of GR as a combination of topical and geographical aspects of relevance. Hence, a simple matching between a category specified in a user query and the category to which a geographic entity belongs to is not a suitable approach.

In this dissertation a semantic distance function will be taken into account as the basis for computing the criterion topicality. The semantic distance function is an inverse measure of the semantic similarity between the user query and the category to which an entity belongs. The higher the similarity, the shorter the semantic distance between the user's information need and a geographic entity belonging to that category. A survey of the literature on methods for measuring semantic relatedness is presented by Zhang et al. (2012). The description of how I implemented such a semantic distance function

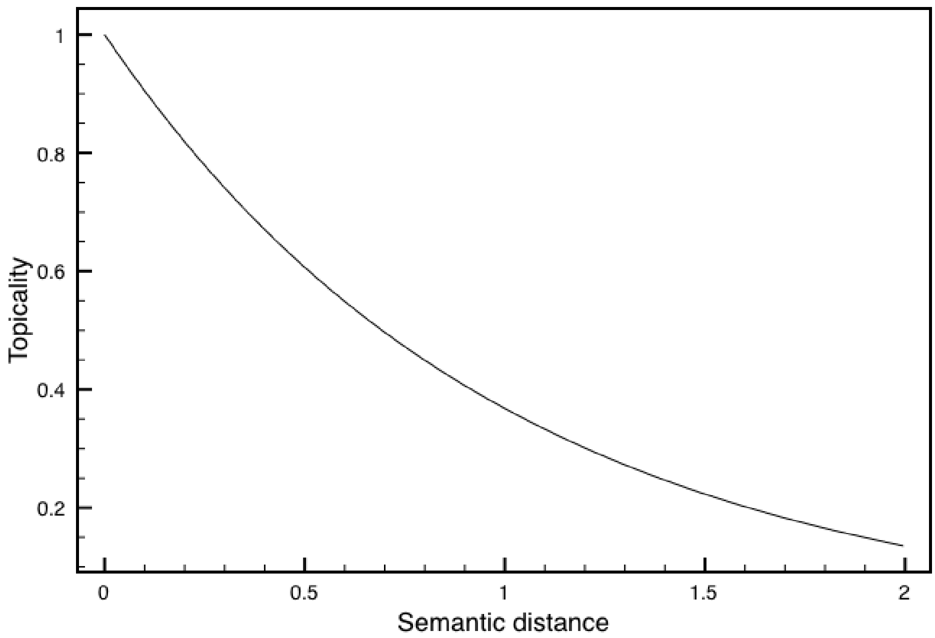


Figure 5.2: Representation of the function used to derive the auxiliary function $d_{Topicality}$ from a given similarity distance between a user query and the category to which the object taken into account belongs.

is reported in Chapter 6.2. In the scope of the current definition, the existence of such a semantic similarity distance function $\sigma(q, g)$ will be assumed as given.

The function $\delta_{Topicality}$ and $d_{Topicality}$ take into account the semantic distance to calculate the strength of the relationship between a user query and a geographic entity for the criterion topicality as shown in Equations 5.3 and 5.4. These formulas use the function $y = e^{(-\lambda \cdot x)}$ (see Figure 5.2) to transform a distance in the range $[0, +\infty]$ into a similarity value in the range $[0, 1]$, where λ regulates the steepness of the decrease. The score for the criterion topicality $\overline{s}_{Topicality}$ is defined as reported in Equation 5.5.

$$\delta_{Topicality}(q, g) = 1.0 - e^{(-\lambda \cdot \sigma(q,g))} \tag{5.3}$$

$$d_{Topicality}(q, g) = e^{(-\lambda \cdot \sigma(q,g))} \tag{5.4}$$

$$\overline{s}_{Topicality}(q, g) = \frac{d_{Topicality}(q, g)}{max_{j \in G}(d_{Topicality}(q, j))} \tag{5.5}$$

5.3.2 Spatio-temporal proximity

The criterion spatio-temporal proximity estimates the proximity between a geographic entity and the user in space and time. This can also be understood as the distance (or difference) between a geographic entity and an hypothetical entity that would perfectly match a user's need, where a perfect match means an entity co-located with a user, and available for as long as a user needs it. This is an inverse measure of utility. Assuming that the user is at a given location, and willing to be at a given destination by a given time, then a perfect match would be at any location along her travel trajectory.

The chosen approach is to compare the amount of time the user needs to perform her activity with the amount of time that she will be able to spend at the location of the entity, considering the travel time needed to reach it and be able to arrive at the destination on time, considering the temporal availability of the entity. The calculations are based on the space-time prism concept (Miller and Bridwell, 2009), as already suggested in the LBS field by Raubal et al. (2004). A similar approach has also been adopted by Pombinho et al. (2012).

The δ_{STprox} distance function takes into account a user's position, the location of an entity, a defined walking speed (i.e., $1ms^{-1}$, that is 3.6 kilometres per hour), a user schedule (i.e., a destination with a mandatory arrival time), a defined minimum time needed to accomplish the activity, and the time validity of the entity (i.e., opening hours). The distance function is then calculated as the ratio between the time needed to fulfil the activity, and the time a user is able to spend at the location of an entity, while the entity is also available (see Equation 5.6). This distance function grows as the amount of time available decreases. If no specification is provided, a forfeit value is taken into account as minimum time needed to accomplish the activity.

The following assumption is also considered: utility grows less than linearly, as the distance value decreases. That is, if an entity is available for twice as long as the user needs, the distance value is cut by half, but the entity is not twice as useful. Thus, the auxiliary function d_{STprox} is defined as a square root function of the inverse of the distance (see Equation 5.7). If the entity is available for less than the time specified by the user as necessary to perform the activity, the utility is zero. The score for the criterion spatio-temporal proximity is defined as shown in Equation 5.8.

$$\delta_{STprox}(q,g) = \frac{time\ needed}{time\ available} \tag{5.6}$$

$$d_{STprox}(q,g) = \begin{cases} 0.0 & \text{if } \delta_{STprox} > 1 \\ \sqrt{\dfrac{1}{\delta_{STprox}}} & \text{otherwise} \end{cases} \tag{5.7}$$

$$\overline{s}_{STprox}(q,g) = \frac{d_{STprox}(q,g)}{max_{j \in G}(d_{STprox}(q,j))} \tag{5.8}$$

It should be noted that the functions proposed in Equations 5.6, 5.7 and 5.8 could produce undesirable results when implemented in a real-world application. On the one hand, temporally unavailable entities should be filtered out, in order not to cause division by zero in Equation 5.6. On the other hand, if no temporal constraint is expressed by the user, entities with non-finite temporal availability (e.g., shops open 24/7, or hotels) would cause division by infinite in Equation 5.6, or division by zero in Equation 5.7. Moreover, assigning an infinite spatio-temporal proximity score to such entities would cause the normalisation procedure in Equation 5.8 to produce undesirable results. A division by infinity would produce either an error, or the assignment of a normalised score equal to zero for all the entities, except those with non-finite temporal availability. Although it seems difficult to imagine a scenario in which the user has no temporal constraint, a real-world application might not be aware of such constraints (e.g., because the user is in a hurry and does not specify them). Even if a user's temporal constraints are taken into account, entities with non-finite temporal availability can produce high local maximum values, which would particularly affect the normalisation step. Therefore, a specific handling of such cases would be necessary for the implementation of a reliable real-world application.

5.3.3 Directionality

The criterion directionality captures the idea that a user would consider those objects which are in the same direction as her destination to be more relevant. As mentioned in Section 5.1, this concept is distinct from the criterion spatio-temporal proximity, but not totally independent. Assuming equal temporal availability, the user will be left with more time to spend at the locations of those entities which are half-way between the current location and destination, than at those entities which are in a different direction. On the one hand, spatio-temporal proximity is strongly related to directionality, if directionality

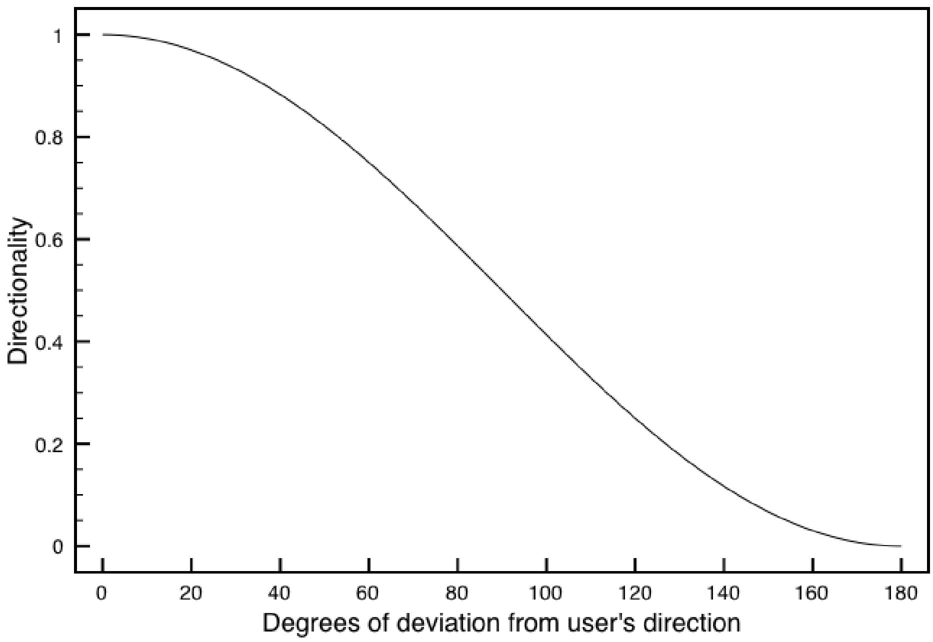


Figure 5.3: Auxiliary function d_{AngDev}, defined for calculating the similarity score for the criterion directionality, given the degrees of deviation between the direction to a user's destination and the direction to a geographic entity.

is understood as a deviation from the user's shortest path to the destination. On the other hand, spatio-temporal proximity is loosely related to directionality, if directionality is understood as the angle between the direction from the current location to the entity's location, and the direction from the current location to the user's destination. An entity near the user's location can have a higher angular deviation than an entity farther away, if it is in the opposite direction with respect to the movement path. In the scope of this dissertation, this second interpretation is taken into account. At the same time, this interpretation demands a careful weighting of the contribution of directionality to the final GR score, so that entities with high spatio-temporal proximity scores are not penalised too much in case of low directionality scores. For instance, a geographic entity in the opposite direction with respect to a user's destination can not be considered totally irrelevant with respect to the mobility component of GR, if it is very close to a user's position, and with a high spatio-temporal proximity score.

The criterion directionality is implemented as a function of the angle between a user's destination and a geographic entity. That is, the angle between a straight line connecting a user's location to the location of her destination, and a straight line connecting the user's location to the location of the entity. The smaller the angle, the lower the value of the distance function δ_{AngDev} (see Equation 5.9), and the higher the value of the auxiliary function d_{AngDev} (see Equation 5.9). The auxiliary function is represented in Figure 5.3, where the resulting value is reported on the y-axis, given the degrees of deviation on the x-axis. The function $cos(\alpha)$ is used in order to obtain a value equal to 1 when the angle α is 0, whereas the distance function δ_{AngDev} returns a value equal to 0 when the angle α is 0 (assuming $0° \leq \alpha \leq 180°$). The resulting value is lower than 0.5 when the angular deviation is higher than 90°. The score for the criterion directionality is calculated as reported in Equation 5.11.

$$\delta_{AngDev}(q, g) = \frac{1 - cos(\alpha)}{2} \tag{5.9}$$

$$d_{AngDev}(q, g) = \frac{cos(\alpha) + 1}{2} \tag{5.10}$$

$$\overline{s}_{Direct}(q, g) = \frac{d_{AngDev}(q, g)}{max_{j \in G} d_{AngDev}(q, j)} \tag{5.11}$$

5.3.4 Cluster

Two complementary aspects will be taken into account to implement the criterion cluster. The first is the size of the cluster (i.e., the cardinality of the cluster) containing the entity under assessment, and captures the idea that a user would prefer a large cluster over a small one. The larger the cluster, the lower is the distance from a hypothetical perfect match for the user's need, the higher the score. The second is the distance between the entity and the closest other entity of the same category, and captures the idea that a user would prefer entities belonging to the same category to be as close as possible to each other. The shorter the distance, the higher the score. The combination of these two aspects conveys information on both the size and the density of the cluster. Moreover, the second aspect entails information about the relationship between an entity and other entities of the same category. This is independent from the parameters used in executing the cluster mining algorithm, which can affect the identification of clusters (e.g., disregarding loose or small clusters), and also results in non-zero scores for those entities that are not in a cluster.

Assuming that the clusters have been mined for the dataset under investigation, the distance function $\delta_{ClustCard}(q, g)$ (see Equation 5.12) is calculated as the ratio between the number of entities in the largest among the clusters $\Phi^{cat(g)}$ of entities belonging to the same category as the entity under assessment g, minus the number of entities in the cluster $\phi(g)$ containing g, and the number of entities in the largest among the clusters $\Phi^{cat(g)}$ of entities belonging to the same category as g. It should be noted that in a real-world application the result of this distance function is influenced by the selection of the area taken into account during the cluster mining process. The distance function $\delta_{ClustDist}(q, g)$ (see Equation 5.13) is calculated as the ratio between the distance between g and the closest entity of the same category, and the distance t_{Clust} used as threshold for mining the clusters. This creates a relationship between the two aspects of the criterion mentioned above. Two auxiliary functions are then calculated as shown in Equations 5.14 and 5.15. The auxiliary function $d_{ClustDist}(q, g)$ employs the same exponential function used for calculating topicality in Equation 5.3, represented in Figure 6.1.

The score for the criterion cluster is calculated as a geometric combination of the two values calculated with the auxiliary functions (see Equations 5.16 and 5.17), as the distance on the Cartesian plane between the origin and the point described by the two values (see Equation 5.16). This approach is a compensatory version of the method used within the SPIRIT Project (Van Kreveld et al., 2005; Purves et al., 2007), and thus it allows a disjunctive combination of the two aspects, which is necessary to obtain non-zero scores for those entities that are not in a cluster, as mentioned above.

There are some categories of entities which tend not to cluster, but instead are almost equally distributed over the geographic space (e.g., pharmacies). In such cases, if no cluster has been identified for a given category, all entities of that category will obtain score 1 for the criterion cluster. The reason for this choice is that, if there are no clusters for a given category, then the criterion is satisfied *a priori*. There is no reason to sustain that in such a case the criterion is not satisfied *a priori* (i.e., assigning the score 0), and other forfeit scores (e.g., assigning the score 0.5) may result in incoherent behaviours when combining this score with others. More sophisticated solutions can be considered in future implementations. For example, the cardinality-based aspect can be given as satisfied and ignored, and the distance-based score can be maintained as sole input to the criterion cluster when no cluster has been identified for a given category.

$$\delta_{ClustCard}(q,g) = \frac{max(\{\|\phi\| \mid \phi \in \Phi^{cat(g)}\}) - \|\phi(g)\|}{max(\{\|\phi\| \mid \phi \in \Phi^{cat(g)}\})} \tag{5.12}$$

$$\delta_{ClustDist}(q,g) = \frac{min(\{dist(g,h) \mid cat(g) = cat(h)\})}{t_{Clust}} \tag{5.13}$$

$$d_{ClustDist}(q,g) = e^{(-\lambda \cdot \delta_{ClustDist})} \tag{5.14}$$

$$d_{ClustCard}(q,g) = 1 - \delta_{ClustCard} \tag{5.15}$$

$$f_{Clust}(q,g) = \frac{\sqrt{d_{ClustDist}(q,g)^2 + d_{ClustCard}(q,g)^2}}{\sqrt{2}} \tag{5.16}$$

$$\overline{s}_{Clust}(q,g) = \begin{cases} 1.0 & \text{if } \Phi^{cat(g)} = \emptyset \\ 0.0 & \text{if } \Phi^{cat(g)} \neq \emptyset \wedge \delta_{ClustCard} = 1 \\ f_{Clust}(q,g) & \text{otherwise} \end{cases} \tag{5.17}$$

5.3.5 Co-location

A co-location rule is composed by a "premises" category and a "conclusion" category. Each rule captures the fact that, given an entity belonging to the first category, it is probable to find an entity belonging to the second category within a pre-defined distance – which has been used as threshold in the mining process. More complex rules, which involves more than one category both as premises and conclusion, are also computable, but these are not taken into account in this dissertation.

Assuming that a set of co-location rules have been mined from the dataset under investigation, a subset of meaningful rules will be taken into account, which can be useful in supporting the user's activity. A rule should be selected only if the affordances related to the "conclusion" category can be considered as correlated with (i.e., subsidiary, complementary, or consequent to) the affordances related to the "premises" category. This would imply a second-order relationship of relevance between the user's query, and the entities belonging to the "conclusion" category.

The definition of the criterion co-location is similar to the one given for the cluster criterion in Section 5.3.4. Given a geographic entity under assessment, the criterion co-location selects the rules in which the category of that entity appears as "premises". For each of these rules, two aspects are considered. The first one is the number of entities within the threshold distance used in the mining process from the entity under assessment, that belong to the "conclusion" category of the rule. The second one is the distance between the entity under assessment, and the closest entity belonging to the "conclusion" category of the rule.

Figure 5.4: Auxiliary function $d_{Coloc^{\psi}Card}$, defined for calculating on the y-axis the co-location cardinality score for hotels, given the number of restaurants within the threshold distance on the x-axis, and assuming 18 as maximum cardinality under consideration.

Thus, the distance function $\delta_{Coloc^{\psi}Dist}(q, g)$ (see Equation 5.20) for the rule $\psi \in \Psi$ is defined as the ratio between the distance between the entity g and the closest entity belonging to the "conclusion" category ψ^c of the rule ψ, and the distance threshold t_{Coloc} used in the mining process. The maximum cardinality of ψ (see Equations 5.18 and 5.19) is defined as the maximum number of entities belonging to the "conclusion" category ψ^c, within the threshold t_{Coloc} used in the mining process from an entity belonging to the "premises" category ψ^p. The distance function $\delta_{Coloc^{\psi}Card}(q, g)$ (see Equation 5.21) for the rule ψ is defined as the ratio between the difference between the maximum cardinality of ψ, and the number of entities belonging to the "conclusion" category ψ^c within the threshold t_{Coloc} used in the mining process, and the maximum cardinality of ψ. It should be noted that in a real-world application, the result of this distance function is influenced by the choice of the area taken into account during the co-location rules mining process.

Two auxiliary functions are calculated as shown in Equations 5.22 and 5.23. The definition of the distance-related similarity function $d_{Coloc^{\psi}Dist}(q, g)$ uses the same exponential function used for calculating topicality in Equation 5.3, and the first cluster score in Equation 5.14, represented in Figure 6.1. The core of the auxiliary function $d_{Coloc^{\psi}Card}(q, g)$ is an inverse function of the value resulting from $\delta_{Coloc^{\psi}Card}(q, g) \in [0, 1]$, which is then decreased by 0.5 and finally multiplied by 2 in order to obtain 1 if $\delta_{Coloc^{\psi}Card}(q, g) = 0$ and the result is 0 if $\delta_{Coloc^{\psi}Card}(q, g) = 1$. An example of the cardinality-related auxiliary function $d_{Coloc^{\psi}Card}(q, g)$ is given in Figure 5.4, for a rule having the category "hotel" as "premises", and the category "restaurant" as "conclusion". The co-location cardinality score for hotels is shown on the y-axis, the number of restaurants within the threshold distance is shown on the x-axis, assuming that the maximum cardinality in the dataset under consideration equals 18 (that is the maximum cardinality obtained for the mentioned rule in one of the test dataset, using a threshold distance of 200 meters).

For each of the co-location rules taken into account in assessing the GR of an entity, a further function $f_{Coloc^{\psi}}(q, g)$ is defined as a disjunctive geometric combination of the values obtained from the auxiliary functions defined above (see Equation 5.24), as it has been issued for the cluster criterion in Section 5.3.4. Finally, the co-location score for the entity under assessment is calculated as the average of the scores related to the different rules, as shown in Equation 5.25. As for the criterion cluster, if no rule has been identified which involve the entity's category as premises, then all the entities of that category are assigned the score 1.

$$card(\psi, x) = \|\{y \mid dist(x, y) \leq t_{Coloc} \wedge cat(x) = \psi^p \wedge cat(y) = \psi^c\}\| \tag{5.18}$$

$$maxCard(\psi) = max(\{card(\psi, x) \mid x \in G \wedge cat(x) = \psi^p\}) \tag{5.19}$$

$$\delta_{Coloc^{\psi}Dist}(q, g) = \frac{min(\{dist(g, h) \mid cat(h) = \psi^c\})}{t_{Coloc}} \tag{5.20}$$

$$\delta_{Coloc^{\psi}Card}(q, g) = \frac{maxCard(\psi) - \|\{h \mid dist(g, h) \leq t_{Coloc} \wedge cat(h) = \psi^c\}\|}{maxCard(\psi)} \tag{5.21}$$

$$d_{Coloc^{\psi}Dist}(q, g) = e^{(-\lambda \cdot \delta_{Coloc^{\psi}Dist})} \tag{5.22}$$

$$d_{Coloc^{\psi}Card}(q, g) = \left(\frac{1}{1 + \delta_{Coloc^{\psi}Card}} - \frac{1}{2}\right) * 2 \tag{5.23}$$

$$f_{Coloc^{\psi}}(q, g) = \frac{\sqrt{d_{Coloc^{\psi}Dist}(q, g)^2 + d_{Coloc^{\psi}Card}(q, g)^2}}{\sqrt{2}} \tag{5.24}$$

$$\overline{s}_{Coloc}(q, g) = \begin{cases} 1.0 & \text{if } \Psi^{cat(g)} = \emptyset \\ \sum_{\psi \in \Psi(g)} \frac{1}{\|\Psi(g)\|} \cdot f_{Coloc^{\psi}}(q, g) & \text{otherwise} \end{cases} \tag{5.25}$$

5.3.6 Example

As an example, assume a user searching for a restaurant for dinner, preferably with pubs nearby where to go for a drink before taking a bus home, as illustrated in Figure 5.5. The minimum time required for dining is 40 minutes. In this example, you want to calculate the scores presented in the previous sections for the restaurant represented in Figure 5.5 with a big dark-orange symbol. Because this entity under assessment is a restaurant (i.e., it is exactly what the user is searching for), the semantic distance (to the query category) is zero. The score for the criterion topicality can be calculated as in Equation 5.26 (using $\lambda = 1.0$). The maximum value for $d_{Topicality}$ is 1.0, as there exists at least one restaurant in the dataset. Since the semantic distance is zero, the resulting score for the criterion topicality is equal to 1.0.

$$\begin{aligned}
\sigma &= \mathbf{0.0} \\
\delta_{Topicality}(q,g) &= 1.0 - e^{0.0} = 1.0 - 1.0 = 0.0 \\
d_{Topicality}(q,g) &= e^{0.0} = 1.0 \\
\overline{s}_{Topicality}(q,g) &= \frac{1.0}{1.0} = \mathbf{1.0}
\end{aligned} \tag{5.26}$$

The user has 80 minutes left for dinner in the restaurant under assessment. There is another restaurant where the user would have 125 minutes left for dinner (therefore

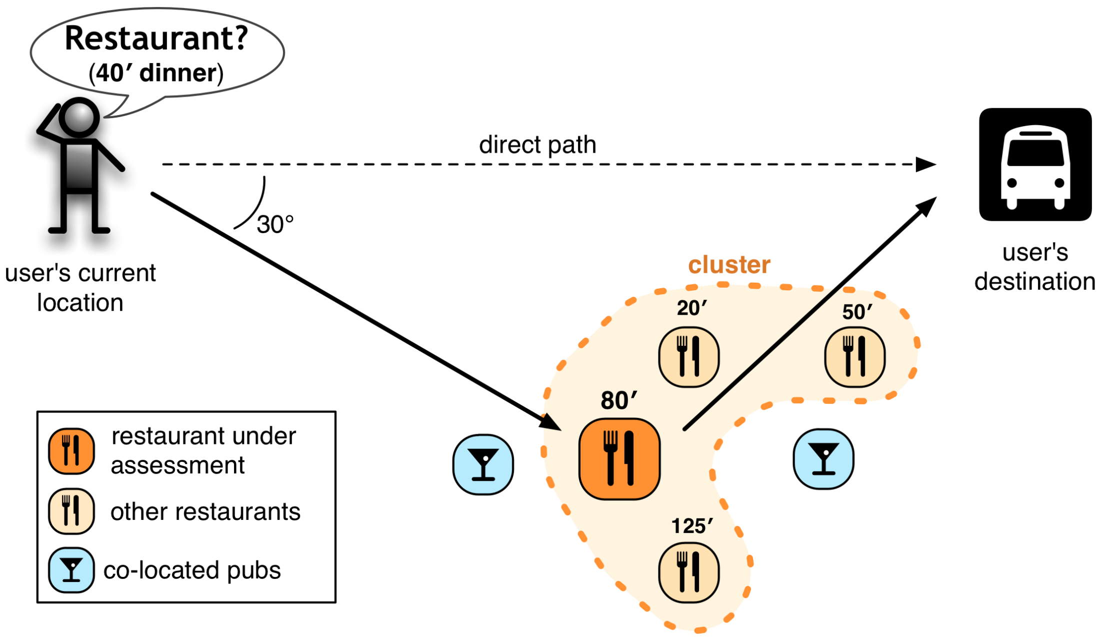


Figure 5.5: Illustration of the situation presented in Section 5.3.6 as an example of GR assessment.

the maximum value for d_{STprox} is 1.76). The score for the criterion spatio-temporal proximity can be calculated as in Equation 5.27. The resulting score for the criterion spatio-temporal proximity is then 0.80.

$$
\begin{aligned}
\delta_{STprox}(q,g) &= \frac{40}{\mathbf{80}} = 0.5 & d_{STprox}(q,g) &= \sqrt{\frac{1}{0.5}} = 1.41 \\
\text{max-}\delta_{STprox}(q,g) &= \frac{40}{\mathbf{125}} = 0.32 & \text{max-}d_{STprox}(q,g) &= \sqrt{\frac{1}{0.32}} = 1.76 \\
\bar{s}_{STprox}(q,g) &= \frac{1.41}{1.76} = \mathbf{0.80}
\end{aligned} \tag{5.27}
$$

The angular deviation from the straight direction to the user's destination is equal to 30°. There is another restaurant at an angular deviation equal to 15°(therefore the maximum value for d_{STprox} is 0.98). The score for the criterion directionality can be calculated as in Equation 5.28. The resulting score for the criterion directionality is then 0.95.

$$
\begin{aligned}
\delta_{AngDev}(q,g) &= \frac{1 - cos(\mathbf{30°})}{2} = 0.07 & d_{AngDev}(q,g) &= \frac{cos(\mathbf{30°}) + 1}{2} = 0.93 \\
\text{max-}\delta_{AngDev}(q,g) &= \frac{1 - cos(\mathbf{15°})}{2} = 0.02 & \text{max-}d_{AngDev}(q,g) &= \frac{cos(\mathbf{15°}) + 1}{2} = 0.98 \\
\bar{s}_{Direct}(q,g) &= \frac{0.93}{0.98} = \mathbf{0.95}
\end{aligned} \tag{5.28}
$$

Assume that the closest other restaurant is at a distance of 80 meters, and that a threshold equal to 100 meters has been used to mine the clusters from the dataset. There are 3 other restaurant in the cluster, as shown in Figure 5.5, leading a total of 4 restaurant. Assuming that the larger cluster found in the dataset contains 5 restaurants, the score for the criterion cluster can be calculated as in Equation 5.29. The resulting score for the criterion cluster is then 0.74.

$$
\begin{aligned}
\delta_{ClustDist}(q,g) &= \frac{\mathbf{80}}{100} = 0.8 & d_{ClustDist}(q,g) &= e^{(-0.5 \cdot 0.8)} = 0.67 \\
\delta_{ClustCard}(q,g) &= \frac{5 - \mathbf{4}}{5} = 0.2 & d_{ClustCard}(q,g) &= 1 - 0.2 = 0.8 \\
d_{ClustCard}(q,g) &= 1 - 0.2 = 0.8 \\
\bar{s}_{Clust}(q,g) &= f_{Clust}(q,g) = \frac{\sqrt{0.67^2 + 0.8^2}}{\sqrt{2}} = \mathbf{0.74}
\end{aligned} \tag{5.29}
$$

Assume that the closest pub is at 100 meters , and that a threshold equal to 200 meters has been used to mine the co-location rules from the dataset, and using $\lambda = 1.0$. There are 2 pubs nearby the restaurant. Assuming that the maximum of pubs co-located with a given restaurant in the dataset is 2, the score for the criterion cluster can be calculated as in Equation 5.30. The resulting score for the criterion co-location is 0.82.

$$\begin{aligned}
\delta_{Coloc^{Pub}Dist}(q,g) &= \frac{\mathbf{100}}{200} = 0.5 \qquad & d_{Coloc^{Pub}Dist}(q,g) = e^{(-1.0 \cdot 0.5)} = 0.60 \\
\delta_{Coloc^{Pub}Card}(q,g) &= \frac{2-\mathbf{2}}{2} = 0.0 \qquad & d_{Coloc^{Pub}Card}(q,g) = (1-0.5)*2 = 1.0 \\
\overline{s}_{Coloc}(q,g) = f_{Coloc^{Pub}}(q,g) &= \frac{\sqrt{0.60^2+1.0^2}}{\sqrt{2}} = \mathbf{0.82} &
\end{aligned} \tag{5.30}$$

5.4 Probabilistic scores

The scores proposed in the previous sections are calculated as functions of the distance values δ. These scores do not take into account the distribution of these values. Hence, for each criterion, an entity assigned with a given distance value δ would obtain the same score $\overline{s}$, independently of the number of other entities with lower distance values (i.e., better fitting the user's need with respect to that criterion). For instance, applying this principle to the sole spatial dimension, an entity would obtain the same score, given its distance from a user's position, no matter whether it is the closest entity to the user or if there are a large number of closer entities.

In De Sabbata and Reichenbacher (2010), I proposed the GRBM25 model, based on the Okapi BM25 model (Spärck Jones et al., 2000), to investigate this aspect of the assessment of GR. The GRBM25 model calculates the probability of an entity being relevant for a given criterion, based on the distance values δ, and their distribution. The aim is to achieve an higher sensitivity to small changes in the user context, and therefore a better understanding of GR. For example, assuming that most clusters of a given category contain three entities, an entity would be assigned with a high probability of being relevant for the criterion cluster if it is a member of a cluster containing five entities. If most clusters of the same category would contain seven entities, that same entity would be assigned with a low probability of being relevant for the criterion cluster. The same applies to all the criteria taken into account.

The definitions proposed in (De Sabbata and Reichenbacher, 2010) are reported below. Equation 5.31 shows how a first value is computed for a given criterion $\gamma \in \Gamma$, assuming that both the distance function δ and the auxiliary function d have been defined. Two auxiliary functions are reported in Equations 5.33 and 5.32 The function originally named *odf* in (De Sabbata and Reichenbacher, 2010) has been renamed ef, which stands for "entity frequency", corresponding to the "document frequency" taken into account in the Okapi BM25 model (Spärck Jones et al., 2000). The first computes the average distance value for a given criterion, while the latter accounts for the number of entities having a distance value equal to or less than the distance value of the given entity. The variables k_1 and b are tuning parameters derived from the original Okapi BM25 formula (see Spärck Jones et al., 2000). The combined probabilities for the criteria cluster and co-location can be calculated as shown in Equations 5.34 and 5.35, where the function *min* refers to the minimum of the two parameters, and corresponds to the fuzzy logic conjunction of the two parameters.

Finally, the calculated P values have to be normalised to obtain the probability values $\overline{P}$ within the range $[0,1]$. For example, P values can be normalised with respect to the maximum value obtained for each criterion. For the criteria cluster and co-location it is preferable to calculate the probability values $\overline{P}$, and to normalise them per category, as each category has its specific distribution of values. For instance, it would be at least questionable to calculate the probability of relevance by comparing the cardinality of the cluster of hotels with the distribution of all clusters, including clusters of restaurants and bars, which are usually found in much larger clusters than hotels.

$$\forall\, \gamma \in \Gamma, \quad P_\gamma(q,g) = log\left[\frac{\|G\|+1}{ef(\delta_\gamma,c,g)}\right] \cdot \frac{(k_1+1)\cdot d_\gamma(q,g)}{k_1\left((1-b)+b\left(\frac{\delta_\gamma(q,g)}{avg(\delta_\gamma,c,G)}\right)\right)+d_\gamma(q,g)} \tag{5.31}$$

$$ef(\delta_\gamma,c,g) = \|\{h \in G | \delta_\gamma(q,h) \le \delta_\gamma(q,g)\}\| \tag{5.32}$$

$$avg(\delta_\gamma,c,G) = \frac{1}{\|G\|}\sum_{g\in G}\delta_\gamma(q,g) \tag{5.33}$$

$$P_{Clust}(q,g) = \begin{cases} 1.0 & \text{if } \Phi(t(g)) = \emptyset \\ 0.0 & \text{if } \Phi(t(g)) \neq \emptyset \wedge \delta_{ClustCard} = 1 \\ min\big[P_{ClustDist}(q,g), \\ \quad P_{ClustCard}(q,g)\big] & \text{otherwise} \end{cases} \tag{5.34}$$

$$P_{Coloc^\psi}(q,g) = \begin{cases} 1.0 & \text{if } \Psi(t(g)) = \emptyset \\ min_{\psi\in\Psi(g)}\big[min(P_{Coloc^\psi Dist}(q,g), \\ \quad P_{Coloc^\psi Card}(q,g))\big] & \text{otherwise} \end{cases} \tag{5.35}$$

5.5 Scores' Combination

This section discusses various methods to combine the five scores proposed in Section 5.3 according to the computational assessment model suggested in Section 5.1 (see Figure 5.1), in order to provide an exploratory answer to the fourth research question (see RQ4, Section 1.2). The proposed approach is evaluated in Chapter 7.

Assuming that the user is searching for an entity that satisfies all the selected criteria as far as possible, the simplest approach is to combine the normalised scores using the arithmetic product (and the equivalent fuzzy logic operator $\wedge$ for the probabilistic scores) to combine the scores. Unfortunately, these methods have a non-compensatory nature; that is, one low score is sufficient to obtain a low aggregate score. This would not be appropriate to combine the geographic environment component (cluster and co-location) with topicality and the mobility component (spatio-temporal proximity and directionality). This combination could produce false irrelevant cases, as the strong "and-ness" of the combination would cause possibly relevant entities to be scored as absolutely irrelevant. For instance, entities would obtain a final relevance score close to zero even if they obtain medium or high topicality and mobility scores, if assigned with a very low geographic environment score. This property is not desirable, because such entities cannot be considered as absolutely irrelevant. Moreover, these methods cause a nonlinear distortion, which is undesirable in most cases. For example, an entity obtaining the maximum normalised score of 1.0 for all five criteria, would be considered 32 times more relevant than an entity obtaining a normalised score of 0.5 for all five criteria (i.e., $1.0^5 = 1.0$ and $0.5^5 = 0.03125$).

Assuming a disjunctive approach to combining the scores, the arithmetic sum (and the fuzzy logic operator $\vee$ for the probabilistic scores) could be used instead. The geometric combination method used in the SPIRIT Project (Van Kreveld et al., 2005; Purves et al., 2007) could also be adapted to account for more than two scores to combine. These methods have a compensatory nature; that is, one high score is sufficient to obtain a medium or high aggregate score. Although this behaviour can be appropriate when all the scores to be combined have the same importance, this is not the case for GR, where the topicality and the mobility components (i.e., spatio-temporal proximity and directionality) are more important than the geographic environment component (i.e., cluster and co-location). This would result in an overestimation of the importance of the geographic environment component, entailing the risk of false relevant cases, and causing topically non-relevant entities to be ranked among the top results.

In order to overcome these issues, the Continuous Preference Logic (CPL) model (Dujmovic, 1975, 2007) is used. This is a continuous logic of decision models, based on the generalised conjunction/disjunction (GCD) function (Dujmovic and Larsen, 2007). The latter allows for the creation of logic operators with any grade of "and-ness" in the range $\alpha \in [0, 1]$ and "or-ness" in the range $\omega \in [0, 1]$, where the sum of the two is always equal to 1. A full "and-ness" operator corresponds to the fuzzy logic operator $\wedge$. A full "or-ness" operator corresponds to the fuzzy logic operator $\vee$. If "and-ness" is higher than "or-ness", the operator is a partial conjunction $\triangle_\alpha$. If "or-ness" is higher than "and-ness", the operator is a partial disjunction $\bigtriangledown_\omega$. The operator characterised by the same grade of "and-ness" and "or-ness" is the arithmetic mean.

CPL builds on such operators to define the conjunctive partial absorption (CPA) and the disjunctive partial absorption (DPA) operators. The CPA operator allows to combine "mandatory" input with "desired" input in a conjunctive manner (see Equation 5.36). The input defined as "mandatory" is accounted as starting value, which is then incremented or decremented, depending on whether the input defined as "desired" is greater or lower than the "mandatory" input, and on the "and-ness" of the partial conjunction and the "or-ness" of the partial disjunction used. If the "mandatory" input is zero, the output will always be zero. Similarly, the DPA operator allows to combine "sufficient" input with "desired" input in a disjunctive manner (see Equation 5.37). The input defined as "sufficient" is accounted as starting value, which is then incremented or decremented, depending on whether the input defined as "desired" is greater or less than the "sufficient" input, and on the "or-ness" of the partial disjunction and the "and-ness" of the partial conjunction used. If the "sufficient" input is equal to 1, the output is always equal to 1. More complex operators can be created combining more GCD functions and the CPA and DPA operators (see Dujmovic, 2007).

$$CPA_{\alpha\omega}(x_{\text{mandatory}}, y_{\text{desired}}) = x_{\text{mandatory}} \bigtriangleup_{\alpha} (x_{\text{mandatory}} \bigtriangledown_{\omega} y_{\text{desired}}) \tag{5.36}$$

$$DPA_{\omega\alpha}(x_{\text{sufficient}}, y_{\text{desired}}) = x_{\text{sufficient}} \bigtriangledown_{\omega} (x_{\text{sufficient}} \bigtriangleup_{\alpha} y_{\text{desired}}) \tag{5.37}$$

I define the score combination for the assessment of GR as illustrated in Figure 5.6. This definition is based on the importance of the criteria discussed in Chapter 4, and takes advantage of the CPL model and GCD functions. Starting from the left-most side, the "distances" calculated by means of the δ functions defined in Section 5.3 are taken as input to calculate the normalised scores $\overline{s}$ (or the probabilistic scores P). The scores are thereafter taken as input by the CPL-based scores combination module. The mobility component is calculated using a CPA operator, taking into account spatio-temporal proximity as "mandatory" input, and directionality as "desired" input (see Equation 5.38). The geographic environment component is calculated as a partial conjunction of cluster and co-location (see Equation 5.39). The resulting value will lie between the minimum and the average of the two input values, depending on the chosen "and-ness". Finally, a CPA operator takes into account topicality and the mobility component as "mandatory" input, including the geographic environment as "desired" input (see Equation 5.40), to return the estimated GR as an output score (see right-most side of the Figure 5.6).

$$\overline{s}_{Mobility}(q,g) = CPA_{0.75\,0.75}(\ \overline{s}_{STprox}(q,g),\ \overline{s}_{Direct}(q,g)\ ) \tag{5.38}$$

$$\overline{s}_{GeoEnv}(q,g) = \overline{s}_{Clust}(q,g)\ \bigtriangleup_{0.75}\ \overline{s}_{Coloc}(q,g) \tag{5.39}$$

$$GR(q,g) = CPA_{0.75\,0.75}(\ \{\ \overline{s}_{Topicality}(q,g),\ \overline{s}_{Mobility}(q,g)\ \},\ \overline{s}_{GeoEnv}(q,g)\ ) \tag{5.40}$$

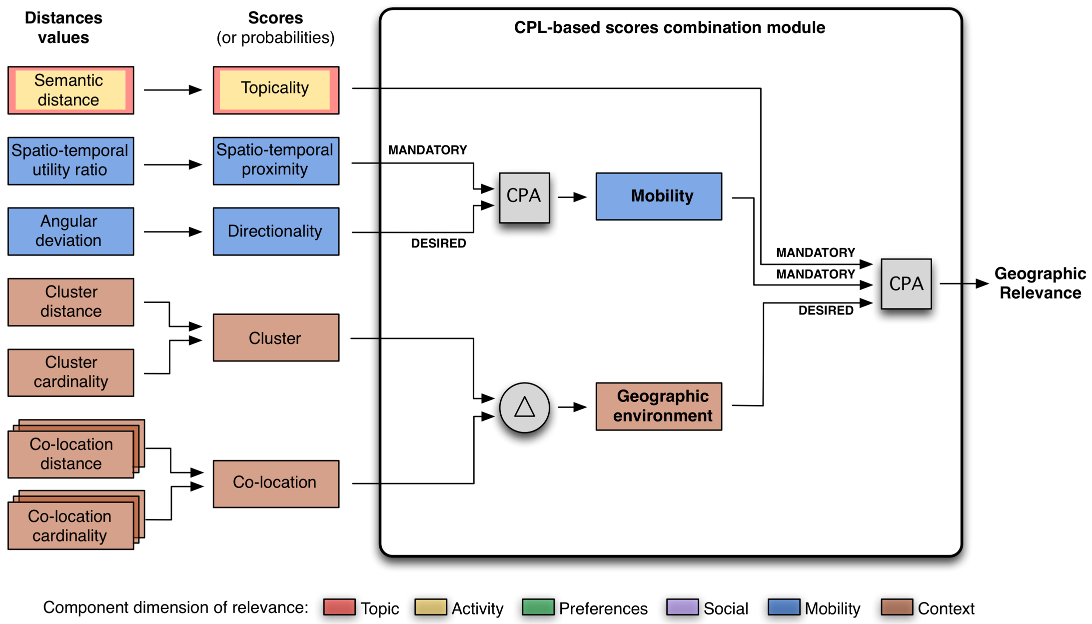


Figure 5.6: Illustration of the GR assessment method based on the CPL model and GCD functions.

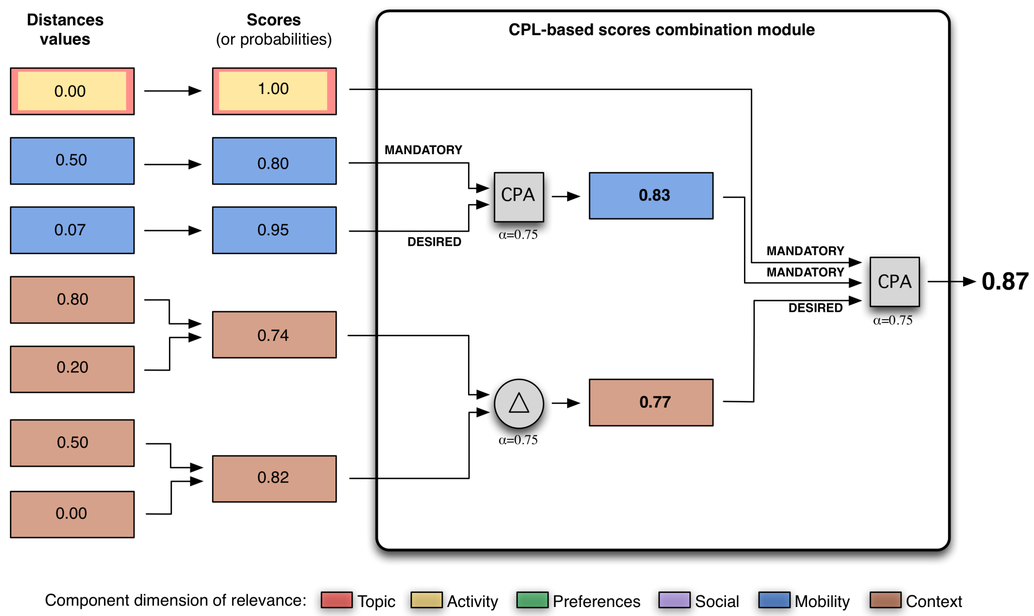


Figure 5.7: Illustration of the GR assessment for the example presented in Section 5.3.6.

This definition is applied to both the normalised scores and the probabilistic scores derived in Section 5.4. Depending on the input, the output of the scores combination module will be the score-based assessment (henceforward referred to as *ScoreGR*) or the probabilistic estimation of GR (henceforward referred to as *GRBM25*). Figure 5.7 illustrates the application of the described method to the scores calculated for the example presented in Section 5.3.6. Using a medium value of "and-ness" and a medium value of "or-ness" (i.e., $\alpha = 0.75$ and $\omega = 0.75$ as described by Dujmovic, 2007), the score related to the mobility component is calculated as equal to 0.83, and the score related to the geographic environment component is calculated as equal to 0.77. These scores are combined to the score related to the criterion topicality (still using a medium value of "and-ness"), resulting in a final score of GR equal to 0.87 (see Equation 5.41).

$$\begin{aligned} GR(q,g) = 0.87 = CPA_{0.75\,0.75}(\ \{\ 1.0,& \\ & CPA_{0.75\,0.75}(\ 0.80,\ 0.95\ )\ \}, \\ & 0.74\ \triangle_{0.75}\ 0.82\ ) \end{aligned} \tag{5.41}$$

5.6 Summary

In this chapter a GR assessment model was presented, which is derived from the conceptual model presented in Chapter 3, and entails the criteria topicality, spatio-temporal proximity, directionality, cluster, and co-location. I identified this as the minimal set including the fundamental and distinguishing criteria of GR, based on the results discussed in Chapter 4. An exploratory answer to the third and fourth research questions (RQ3 and RQ4) is provided, issuing a formal definition of the scores related to each criterion, and proposing a schema to combine them into an aggregate estimation of GR, based on the CPL model and GCD functions.

In order to evaluate the proposed approach, the GR assessment model was prototypically implemented, including the algorithms used to mine clusters and co-location rules. Chapter 6 reports the implementation of the prototype, and Chapter 7 discusses the results of the evaluation.

Chapter 6

Prototype implementation

In order to evaluate the GR assessment methods proposed in Chapter 5, I implemented a prototype GR assessment service software. The objective is to perform the assessment procedure on real data in realistic scenarios, and to compare the calculated results with actual human relevance judgements, in order to evaluate the effectiveness and validity of the proposed approach. The results of this experimental study are reported in Chapter 7. The following sections report the implementation choices made in developing the GR assessment prototype. The software has been implemented in Java[1] Standard Edition 1.6, using the Eclipse[2] 3.7 development environment, on a Apple Mac Pro[3] (2 x 2.66 GHz Dual-Core Intel Xeon with 8 GB 667 MHz DDR2 RAM) running Mac OS X Lion[4] 10.7.4 operating system. The data are stored in PostgreSQL[5] and PostGIS[6] databases, which have been set up on a local-network, virtual server running Ubuntu[7] 10.04 operating system.

6.1 Data

Different data sources were used during the development of the prototype GR assessment service, depending on the current necessities and data availability. OpenStreetMap[8] was one of the main data sources for the overall GeoRel project, as it offers the possibility to download geographic data under the Open Data Commons Open Database License[9] and Creative Commons Attribution-ShareAlike 2.0[10] licences, and with no fee. Recent studies have found OpenStreetMap to provide good quality data (Haklay, 2010), although possibly volatile (Mooney and Corcoran, 2012). However, this is not among the main concerns of this dissertation, as the evaluation does not include any comparisons

[1] `http://www.java.com/`, last accessed July 2012.
[2] `http://www.eclipse.org/`, last accessed July 2012.
[3] `http://www.apple.com/macpro/`, last accessed July 2012.
[4] `http://www.apple.com/osx/`, last accessed July 2012.
[5] `http://www.postgresql.org/`, last accessed July 2012.
[6] `http://postgis.refractions.net/database`, last accessed July 2012.
[7] `http://www.ubuntu.com/`, last accessed July 2012.
[8] `http://www.openstreetmap.org/`, last accessed April 2012.
[9] `http://opendatacommons.org/licenses/odbl/`, last accessed November 2012.
[10] `http://creativecommons.org/licenses/by-sa/2.0/`, last accessed November 2012.

between the information presented to the user and the real world. On the contrary, for experimental reasons, measures have been taken to produce maps that are unfamiliar to test participants, to avoid any potential confounding variable.

The city of Madrid (Spain) was chosen for the evaluation (see Appendix C for further details). The related OpenStreetMap data was downloaded from teczno.com[11] (see OpenStreetMap Wiki[12]) on December 4th, 2011. The software tool Osmosis[13] was used to extract map features[14] with the following tags: leisure, amenity, shop, tourism, historic. The obtained dataset was further processed by deleting minor categories (e.g., "bench", "fire_hydrant", or "flower") which would not have been involved in the testing scenarios as reasonable answer to the user's information need, nor in the mining of co-location rules. The main objective of this pre-processing step was to shorten the computation time of the data mining and relevance assessment procedures. In order to avoid the possibility for the participants to recognise the place, a rotation was applied to the data (60°clockwise or 105°counterclockwise). The maps were presented at a large scale (approximately 1 : 10000) and without labels. Further details on the data and maps used in the empirical evaluation are reported in Appendix C.

6.2 Topicality

In order to implement the criterion topicality as discussed in Section 5.3.1, it is necessary to define a semantic distance measure. As a development of a complete ontology is out of the scope of this thesis, one possible option would have been to link both the geographic entity categories in the dataset and the terms in a given user query to an existing general ontology. Projects such as ConceptNet[15] (Liu and Singh, 2004) and WordNet[16] (Miller et al., 1995) have been tested for this purpose, as they are frequently used for estimating the similarity between categories of points of interest (e.g., Yu et al., 2003; Laukkanen et al., 2004; Lieberman et al., 2004; Skouteli et al., 2005). For instance, D'Ulizia et al. (2010) suggested to use this approach, along with the approach developed by Lin (1998), to match user's query and entities' categories in LBS.

However, a preliminary test on the dataset described above suggested that this approach is less effective for estimating the similarity between a category and a general user query. That is, it can effectively measure that a bar is more similar to a cafe than to a bank, but it did not perform well enough for the estimation of the similarity between a given category, and the activity "shopping" or "having lunch". For example, the concept-network path in ConceptNet between "lunch" and "fountain" through "drink" can be as long as the path between "lunch" and "restaurant" through "eat". This would result in equal topicality score for those two categories "fountain" and "restaurant" for

[11] `http://metro.teczno.com/`, last accessed July 2012.
[12] `http://wiki.openstreetmap.org/wiki/Planet.osm`, last accessed July 2012.
[13] `http://wiki.openstreetmap.org/wiki/Osmosis`, last accessed July 2012.
[14] `http://wiki.openstreetmap.org/wiki/Map_features`, last accessed July 2012.
[15] `http://web.media.mit.edu/~hugo/conceptnet/`, last accessed April 2012.
[16] `http://wordnet.princeton.edu/`, last accessed April 2012.

the user's query "lunch". As this is the kind of reasoning that the GeoRel project is aiming for, I decided to explore a more flexible approach to similarity assessment for mobile information services.

Cilibrasi and Vitanyi (2007) argued that the Web can provide a wider and more flexible source of knowledge for computer-based common-sense reasoning. They suggested a measure to compute the similarity between two words, based on the number of pages in the Web containing each word alone, and the number of pages containing both words. This approach has been named Normalized Google Distance (NGD), as its calculation was originally based on the results obtained through the Google Web Search API[17] (deprecated as of November 1st, 2010). The distance value σ between two words x and y is calculated as:

$$\sigma = NGD(x, y) = \frac{max(logf(x), logf(y)) - logf(x, y)}{logN - min(logf(x), logf(y))} \tag{6.1}$$

where $f(x)$ is the number of pages containing x, $f(y)$ is the number of pages containing y, $f(x, y)$ is the number of pages containing both x and y, and N is a normalising factor which has to be greater than both $f(x)$ and $f(y)$.

At the time of writing, Google as well as other web search engines have limitations on the number of free search requests performable with their APIs. The number of search requests needed for each query to the prototype system is proportional to the number of categories, which counts in hundreds. Therefore, also considering the time needed for an API request compared to a local database query, a local solution has been adopted. In particular, I used the ClueWeb09 dataset[18] the form provided by ReVerb[19] (part of the KnowItAll[20] project, see Fader et al., 2011). In order to speed up the querying process, I decomposed the arguments of the tuple elements containing more then one word, and replaced them with simpler tuples. For instance, the tuple ("spaghetti carbonara", "eat", "italian restaurant") was decomposed in four tuples including ("spaghetti", "eat", "restaurant") and ("spaghetti", "eat", "italian"). This simplified version of the dataset tuples were used in the assessment process. I assumed the query to be composed by a single term (as it is the case in the scenarios taken into account in the empirical evaluation, see Chapter 7), and I also reduced the category names to single terms, by means of a manual stemming process (e.g., "fitness station" to "fitness", and "pizza service" to "pizza").

The defined similarity measure (Equation 6.1) was applied to the definitions issued in Section 5.3.1, using $\lambda = 1.0$. Equation 6.2 reports the obtained semantic distance function, calculated for the query q and the category of the geographic entity g. Equation 6.3 reports the related auxiliary function, which is an inverse measure of their semantic similarity defined in Equation 6.2. The graph presented in Figure 6.1 illustrates some examples of the results obtained applying this approach to the data used in the

[17] `https://developers.google.com/web-search/`, last accessed April 2012
[18] `http://lemurproject.org/clueweb09.php/`, last accessed April 2012.
[19] `http://reverb.cs.washington.edu/`, last accessed April 2012.
[20] `http://www.cs.washington.edu/research/knowitall/`, last accessed April 2012.

experiments.

$$\delta_{Topicality}(q, g) = 1.0 - e^{-NGD_{ReVerb}(q, g_{Category})} \tag{6.2}$$

$$d_{Topicality}(q, g) = e^{-NGD_{ReVerb}(q, g_{Category})} \tag{6.3}$$

The method proposed above shares some of the disadvantages of the ontology-based approaches, such as possible misunderstandings of ambiguous words, and lack of completeness. Nevertheless, the manual stemming is faster, and easier to apply to other datasets with respect to category-to-concept matching. In general, this approach seems to be more stable, probably due to the large, and diverse underlying dataset. However, this approach can also produce unsatisfying semantic similarity results for some category pairs. For instance, in the example illustrated in Figure 6.1, pubs are rated almost as similar to hotels as hostels are to hotels.

This issue will become an even more crucial point as soon as similar systems will have to deal with more complex, natural-language queries, and growing amounts of user-generated content, and linked data (Janowicz et al., 2012). At the same time, Alves and Pereira (2012) suggest that the diverse sources available on the web (e.g., Flickr[21] and Wikipedia[22] – see e.g., Pereira et al., 2009) can be combined with upper-level ontologies (e.g., WordNet[23] – see e.g., Alves et al., 2009) to uncover the meaning of a place, in order to produce a tag cloud describing its affordance. Alazzawi et al. (2012) suggested to focus on geography-related corpora (i.e., Ordnance Survey's Real World Object Catalogue[24]) to mine knowledge about affordances usually associated with entity categories, using linguistic analysis, where verb phrases denote service- and activity-related concepts. These methods could be combined with the approach suggested by Mizzaro and Vassena (2011), which employes tag clouds to describe the social context and activity of a user, to achieve a finer matching between the user's information need, and the affordances of the entities.

6.3 Spatio-temporal proximity and directionality

In the prototype GR assessment service software, I developed two Java methods to calculate the spatial distance between two points on the Earth's surface. The first calculates Euclidean distances[25], whereas the second computes route-network distance, both implemented using GeoTools[26] Java library.

In order to compare the two methods listed above, I computed the Euclidean and route-network distances for all the possible pairs of points of interest registered in Open-

[21] `http://www.flickr.com/`, last accessed November 2012.
[22] `http://www.wikipedia.org/`, last accessed November 2012.
[23] `http://wordnet.princeton.edu/`, last accessed November 2012.
[24] `http://www.ordnancesurvey.co.uk/`, last accessed November 2012.
[25] This has been calculated as orthodromic distance, using the method *getOrthodromicDistance* of the class *GeodeticCalculator* included in the GeoTools Java library.
[26] `http://www.geotools.org/`, last accessed May 2012.

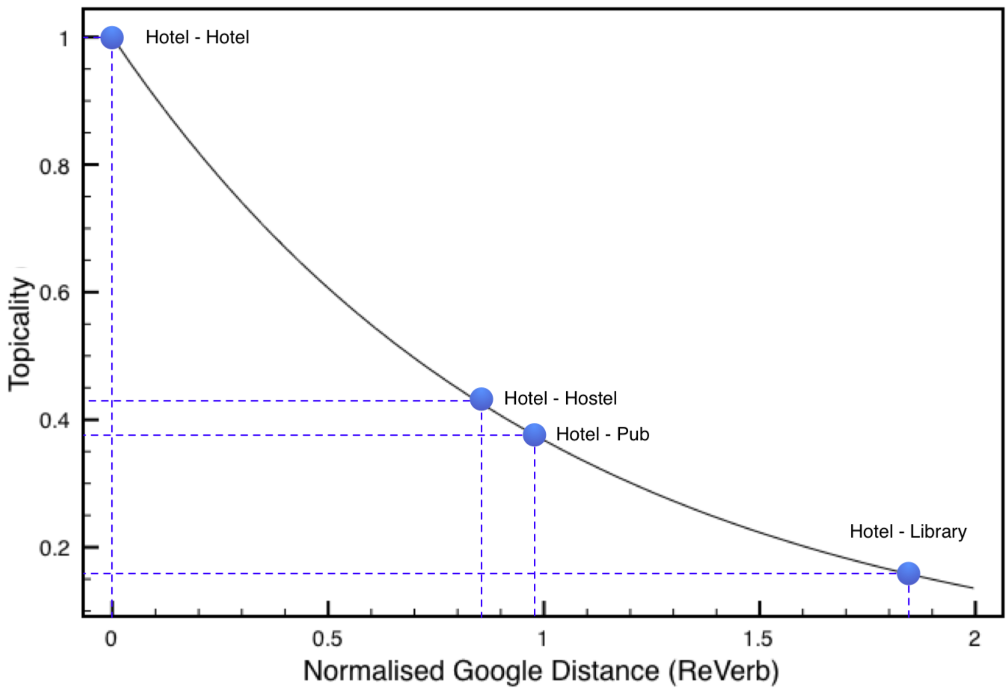


Figure 6.1: Example of the function developed to compute the auxiliary function $d_{Topicality}$ in Equation 6.3, with respect to the user query "hotel".

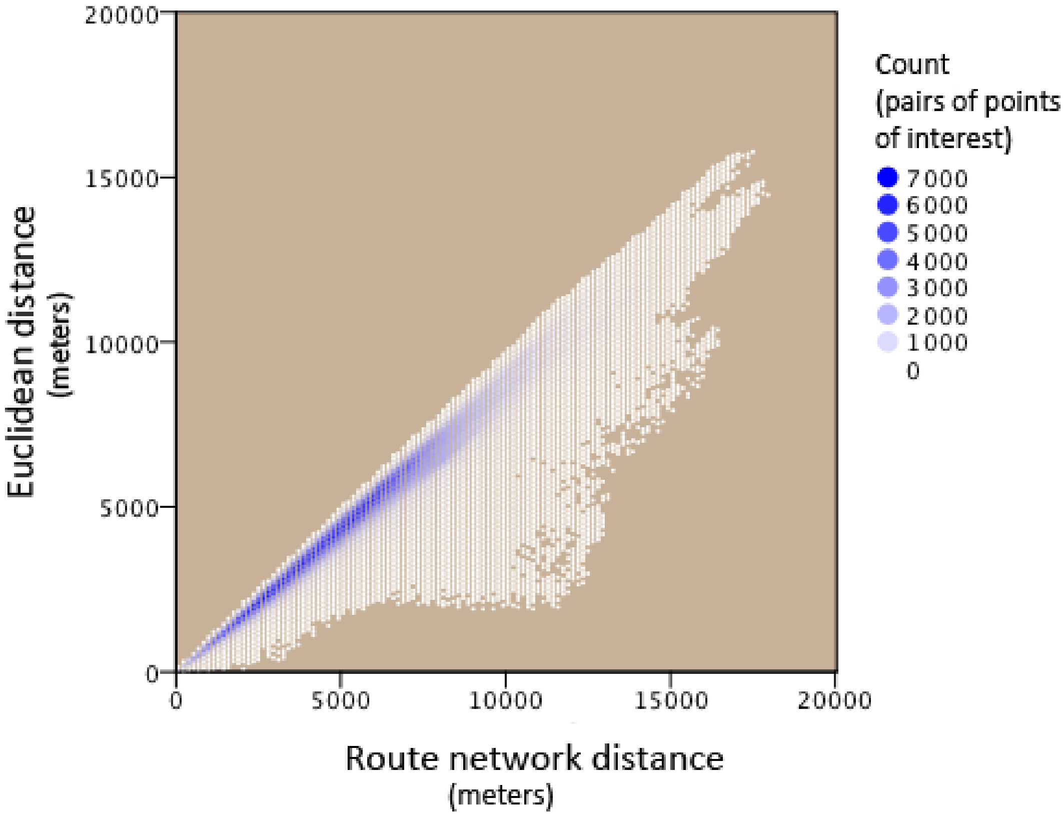


Figure 6.2: Difference between Euclidean and route-network distance calculated for all the possible pairs of points of interest registered on OpenStreetMap in the area of Zürich.

Table 6.1: Pearson's correlation coefficients calculated between Euclidean, Manhattan, and route-network distances.

	r	p
Euclidean and route-network distances	.978	<.01
Manhattan and route-network distances	.954	<.01

StreetMap in the area of Zürich (see Figure 6.2). About 1.8 million pairs were considered, with distances up to about 20 kilometres. On the one hand, the Euclidean distance was on average 23% shorter than the route-network distance, but this average difference drops to 18% when only considering the route-network distances up to 300 metres. These results include many outliers, due to the spatial configuration of the city of Zürich, enclosed between hills and a lake, and divided by two rivers. The Pearson's correlation coefficients calculated between Euclidean and route-network distances, and between Manhattan and route-network distances are reported in Table 6.1. The correlation coefficients show that 95.6% of the variability of route-network distance is accounted for by the Euclidean distance within the considered area. These insights have been confirmed in a recent study by Boscoe et al. (2012), who analysed the differences between Euclidean and route-network distances calculated from locations in residential areas to hospitals. The results of the mentioned study show that the two calculated distances are highly correlated, with important local exceptions near shorelines and other physical barriers, similar to what was reported above concerning the Zürich test.

At the same time, the running time of the developed route-network method is about ten times higher than the running time of the method calculating the Euclidean distance. This would shift the time needed to assess GR for a given scenario from minutes (mostly spent calculating the semantic distance, as described in Section 6.2) to hours. Moreover, the route-network distance is still an approximation of the real distance, affected by missing links and node density, especially when computing walking paths as these networks are optimised for car navigation. Finally, the travel time derived from the spatial distance is a further approximation of reality. Therefore, I decided to base the GR assessment computation on the method calculating the Euclidean distance.

When calculating the score for the criterion directionality, an angle of zero degrees is assigned to all the geographic entities, if the user location and the destination overlap. In such cases, all geographic entities will obtain the maximum score of 1 for the criterion directionality.

6.4 Cluster and co-location

In order to implement the criteria cluster and co-location, I implemented the cluster algorithm GDBSCAN proposed by Sander et al. (1998) and the co-location rule mining algorithm proposed by Huang et al. (2004) (see Section 2.4).

In mining the clusters, I set the cardinality threshold so that a minimum of three entities was required for a cluster to be acknowledged. I set the maximum distance between the objects to 100 metres (i.e., the *eps* variable in the definition given by Ester et al. (1996)), calculated as Euclidean distance. The same value was applied (as divisor) to calculate the distance function $\delta_{ClustDist}(q,g)$. In order to derive the auxiliary value $d_{ClustDist}(q,g)$, I set the parameter to $\lambda = 0.5$, so that an entity having another entity belonging to the same category within the threshold used for clustering (i.e., 100 metres) would obtain a resulting value higher than 0.6 for this specific aspect of the criterion cluster.

In mining the co-location rules, I set the distance threshold to 200 metres (calculated as Euclidean distance), the minimum prevalence threshold to $\theta = 0.33$ (i.e., one third of the entities of a given category have to be involved in the rule), and the conditional probability threshold to $\alpha = 0.33$ (i.e., there has to be a 33% probability of finding an entity belonging to the "conclusion" category nearby a "premises" category). Only rules with one category as premises and one category as conclusion were mined (i.e., the parameter has been set to $k = 2$). Among the 312 discovered rules, only 46 of them were retained I preformed a manual selection, based on the semantic meaningfulness of the rules, considering whether a user searching for an entity belonging to the "premises" category would be interested in an entity belonging to the "conclusion" category. For instance, I dismissed a mined rule, which was conveying the 33% probability of finding a hairdresser nearby one of the three seafood shops included in the mined dataset. In order to derive the auxiliary value $d_{Coloc^{\psi}Dist}(q,g)$, the parameter was set to $\lambda = 1.0$, so that entities having other entities belonging to the "conclusion" category within the threshold used for the mining the co-location rules (i.e., 200 metres) obtain a resulting value higher than 0.3 for this specific aspect of the criterion colocation, and a resulting value higher than 0.6 if the distance is less than half of the threshold used.

6.5 Values combination

The combination of the scores was implemented as described in Section 5.5 (see Figure 5.6), using the continuous preference logic operators described by Dujmovic (2007). In using both the CPA operators and the partial conjunction operators, I selected a medium levels of "and-ness", setting the parameter to $\alpha = 0.75$, consequently $r = 0.\overline{3}$ (see definitions of functions h_α and z, p. 1094, Dujmovic, 2007). I run a few preliminary GR assessments to test slightly different levels of "and-ness", from strong ($\alpha = 0.8750$) to medium-weak ($\alpha = 0.6875$). No noteworthy changes in the final result of the combination were found.

Chapter 7

Evaluation

In this chapter I discuss the effectiveness of the two GR assessment methods proposed in Chapter 5, and implemented as described in Chapter 6. Effectiveness is here understood as the similarity between the rank produced by a GR assessment method and relevance judgements performed by human subjects.

The "user-centred" evaluation of the proposed GR assessment methods follows the common IR benchmark-based evaluation procedures, where a system output is compared to relevance judgements from human subjects. Unfortunately, at the time of writing, no benchmark that could be used to evaluate the effectiveness of the proposed methods was available. As mentioned in Section 2.5, the current Contextual Suggestion Track[1] (part of TREC 2012[2]) has similar perspectives, but is not applicable to GR, due to the adopted description of context and granularity of spatio-temporal information. Thus, I decided to follow the approach employed by Urbano et al. (2010), using a crowdsourcing service (see Section 2.5.1) to collect relevance judgements and create a new benchmark.

7.1 Experiment III

Experiment III was set up to achieve a user-centred evaluation of the GR assessment methods proposed. Section 7.1.1 describes the methods used in Experiment III, with references to Appendix C where the materials used are reported in detail. The obtained results are then presented in Section 7.1.2, and discussed in Section 7.1.3.

7.1.1 Methods

In the following sections, "*ScoreGR*" refers to the main score-based GR assessment method (see Table 7.1), which is based on the scores presented in Section 5.3, and the scores combination method proposed in Section 5.5. The acronym "*GRBM25*" is used to refer to the GR assessment method based on the probabilistic model GRBM25 presented in Section 5.4 and the scores combination method proposed in Section 5.5.

[1] `https://sites.google.com/site/treccontext/`, last accessed September 2012.
[2] `http://trec.nist.gov/pubs/call2012.html`, last accessed September 2012.

Table 7.1: GR assessment methods tested in Experiment III.

Method	criteria	scores	scores combination
Baseline 1	topicality	category-based filter	
	spatial prox.	order by user's path length	
Baseline 2	topicality	see Section 5.3.1	geometric combination (Purves et al., 2007)
	spatial prox.	normalised inverse value of the user's path length	
ScoreGR	topicality spatio-temporal prox. directionality cluster co-location	see Section 5.3	see Section 5.5
GRBM25	topicality spatio-temporal prox. directionality cluster co-location	see Section 5.4	see Section 5.5

Two baseline methods are also considered, in order to compare the effectiveness of simpler assessment models with the effectiveness of the methods under investigation. The first baseline method resembles a very simple LBS approach, and it will be referred to as "*Baseline1*". Given a query, *Baseline1* filters out all entities whose category does not match the query text, and orders the remaining entities according to the length of the user's path (i.e., the distance from the user's current location to the location of the entity, and then to the destination). The second baseline method will be referred to as "*Baseline2*", and takes advantage of the topicality score discussed in Section 5.3.1 (implemented as described in Section 6.2). *Baseline2* combines the topicality score with a distance score computed as the inverse of the length of the user's path (i.e., normalised in the range $[0, 1]$, dividing it by the maximum obtained value), using the geometric combination method adopted in the SPIRIT Project (Van Kreveld et al., 2005; Purves et al., 2007).

Scope. The findings presented in Chapter 4 account for the importance of the criteria implemented by *ScoreGR* and *GRBM25*, i.e., topicality, spatio-temporal proximity, directionality, cluster, and co-location. The aim of Experiment III is to verify whether the proposed GR assessment methods effectively assess GR as a combination of the five selected criteria in the given scenarios. The effectiveness of the proposed methods is measured as their correlation with crowdsourced ranks, which are accounted as "ground truth".

I designed three scenarios, which involve clusters of geographic entities, co-location rules, and spatio-temporally unaccessible entities. In order to perceive those three aspects and consequently influence the output rank, a GR assessment method should account for the criteria related to those three aspects, which are implemented in *ScoreGR* and *GRBM25*, but not in the two baseline methods. Therefore, an additional objective is to establish whether the baseline methods provide a sufficient approximation of GR, even if they do not explicitly implement the criteria spatio-temporally, cluster, and co-location. I did not consider simpler scenarios (e.g., if the user is searching for a type of geographic entity which is not involved in cluster or co-location rules), because in such cases the additional criteria implemented by *ScoreGR* and *GRBM25* do not influence the output rank, by design (see Sections 5.3 and 5.5). Therefore, in such simpler scenarios, *ScoreGR* and *GRBM25* would resemble the output of the baseline methods, apart from the scores combination.

Participants. A total of 416 participants took part in this experiment. The participants were gathered through the online service CrowdFlower[3], which posts the tasks to the crowdsurcing platform Amazon Mechanical Turk[4]. I assumed that the participants fall into typical Amazon Mechanical Turk's demographics (Ross et al., 2010), i.e., computer literate people with no particular expertise in geography. All participants in this experiment were found to connect with a U.S. IP address. Hence, all participants are assumed to be familiar with the simple urban scenarios described in Appendix C, and thus good candidates for this empirical study, which addresses the general public. The allocation of participants to the different scenarios and iterations is presented in Appendix C.

Material. All three scenarios are set in an urban environment, using the dataset described in Section 6.1. In the first scenario, the user is searching for a supermarket while returning home from work. In the second scenario, the user is searching for a hotel nearby where she is attending a conference. In the third scenario, the user is searching for a restaurant. A detailed description of the materials used in this experiment is reported in Appendix C, including instructions and stimuli presented to the participants.

Procedure. For all the three scenarios, the first part of the procedure adopts the "pooling" approach commonly used in IR (see e.g., Manning et al., 2008, p. 151). Due to temporal and budget constraints of the project, it would not be possible to collect relevance judgements for each of the thousands of geographic entities contained in the dataset used in the prototype (see Section 6.1). Therefore, following the "pooling" approach, the same relevance assessment methods under investigation were taken into account to select a smaller sample of entities to judge. The relevance computation and ranking were performed using all four methods, and entities in the top-k list of at least one of the methods have been included in the set of entities to be judged. The underlying assumption is that a relevant geographic entity would be recognised by at least one of the methods. In practice, many of the elements in the top-k lists of the four methods were common to at least two or three methods, which is a strong indication that the relevant entities have been identified. Moreover, a subsequent manual check of the dataset did not identify any clearly relevant entity among the excluded ones. Further details for

[3]`https://crowdflower.com`, last accessed September 2012
[4]`https://www.mturk.com/mturk/welcome`, last accessed September 2012

each scenarios are reported in Appendix C.

The second part of the procedure follows the approach used by Urbano et al. (2010). For each scenario, given an unordered list of entities identified in the previous part of the procedure, one entity is randomly selected[5] as pivot. Thereafter, an iteration is created, as the list of pairwise comparisons between the pivot and all the remaining entities of the list, ordered in a random manner. For each comparison, the labels "A" and "B" are randomly assigned to the two entities, and the order of presentation is randomised. One up to three "check comparisons" are added to each iteration containing three or more comparisons. Check comparisons are a duplicate of one of the comparisons of the iteration, where the order of presentation of the entities, or the "A" and "B" labelling, or both have been swapped. In order to account for a participant learning effect, the Latin Square method has been used to produce different orderings, each following the same order, but starting with a different comparison. The iteration is finally posted on the crowdsourcing service CrowdFlower, with a minimum of 40 participants per iteration, and at least 4 participants per different order. For each comparison, participants were asked to choose between the two entities, and to provide a textual motivation for their judgement. Participants could also choose to classify the two entities under comparison as equally relevant, or as both non-relevant. A detailed description of the procedure and stimuli is presented in Appendix C.

7.1.2 Results

The answers provided by the workers[6] (see Appendix C for further details) were used to generate a rank for each scenario, which will be referred to in the following with the label "*Crowdsourced*". These ranks seem meaningful and coherent with the provided scenarios and instructions.

The *Crowdsourced* ranks are used as "ground truth" and are compared with the ranks generated by *Baseline1*, *Baseline2*, *ScoreGR*, and *GRBM25*. For this purpose, the Kendall's τ correlation coefficient (Kendall, 1938), which is commonly used in statistics to compare ranks in the case of small datasets with a large number of tied elements, is computed. The use of a statistical rank correlation coefficient in IR was first suggested by Pollock (1968), and first applied by Joachims (2002), using the Kendall's τ coefficient, as reported by Sanderson (2010). The results of the statistical calculations[7] for the three scenarios are reported in Tables 7.2, 7.3, and 7.4, respectively.

For all three scenarios, there is a significant correlation ($p < .05$) between the crowdsourced rank and *ScoreGR*, and in the case of Scenario 2 the correlation is highly significant ($p < .01$). The results show that *Crowdsourced* (i.e., the assumed "ground truth" rank) accounts for 30% of the variability of *ScoreGR* in Scenario 1, for 74% in Scenario

[5] Only one exception has been made to the random selection procedure, in Scenario 3 iteration 4, for specific reasons – as reported in Section C.4.

[6] In crowdsourcing, the participants to an experiment are commonly referred to as "workers", since they are paid for their work.

[7] The forfeit rank "999999" has been assigned to entities identified as irrelevant, both in the case of the crowdsourced ranks, and in the case of the four evaluated methods.

Table 7.2: Comparison between crowdsourced and computed ranks for Scenario 1.

Entity	*Crowdsourced*	*Baseline1*	*Baseline2*	*ScoreGR*	*GRBM25*
9128	1	7	7	1	2
9127	2	3	3	4	4
9126	3	5	5	6	7
9124	4	8	8	5	8
9115	5	4	4	2	1
9117	6	2	2	3	3
9125	7	6	6	8	6
9121	8	9	9	7	5
9123	IRR	1	1	IRR	IRR
Correlation	Kendall's τ	-.111	-.111	.556*	.333
	p	>.05	>.05	<.05	>.05

Table 7.3: Comparison between crowdsourced and computed ranks for Scenario 2.

Entity	*Crowdsourced*	*Baseline1*	*Baseline2*	*ScoreGR*	*GRBM25*
9694	1	2	6	1	2
9696	2	5	14	4	5
9700	2	6	16	3	6
9698	4	10	21	2	10
9693	5	3	7	6	3
9828	6	1	2	7	1
9695	7	4	10	8	4
675	IRR	IRR	4	206	51
677	IRR	IRR	1	193	41
5912	IRR	IRR	3	77	40
Correlation	Kendall's τ	.458	-.442	.861**	.442
	p	>.05	>.05	<.01	>.05

Table 7.4: Comparison between crowdsourced and computed ranks for Scenario 3.

Entity	*Crowdsourced*	*Baseline1*	*Baseline2*	*ScoreGR*	*GRBM25*
714	1	1	1	2	1
704	2	5	5	1	3
7212	3	13	13	5	13
7213	3	12	12	4	9
724	5	3	3	38	4
7211	5	19	19	3	20
747	7	2	2	15	2
746	8	7	7	17	5
711	IRR	4	4	IRR	IRR
Correlation	Kendall's τ	.057	.057	.686*	.400
	p	>.05	>.05	<.05	>.05

* Correlation is significant at the 0.05 level (2-tailed).
** Correlation is significant at the 0.01 level (2-tailed).

Note: the label "IRR" refers to entities identified as irrelevant.

2, and for 47% in Scenario 3. No significant correlation was found between *Crowdsourced* and any of the baseline methods in any scenario, nor between *Crowdsourced* and *GRBM25*.

The judgements delineate a clear result for most of the comparisons, with only two exceptions in Scenario 3 (see Appendix C for a detailed description of this scenario). In iteration 2, it is not clear whether the entity 704 is judged as more or equally relevant than the pivot. Therefore, the same entity 704 is taken into account as pivot in iteration 4 (see second and third row in Figure C.10, Appendix C), and compared to all the entities which have been judged more or equally relevant than the previous pivot (i.e., entity 7213). In that iteration, entity 704 was finally judged more relevant than entity 7213. Therefore iteration 4 settled the uncertainty that arose in iteration 2, and the conclusions drawn in the next section are not affected by that uncertain result.

The second exception is the comparison between the entities 746 and 711 in iteration 1 (see first row in Figure C.10, Appendix C). Entity 711 refers to a restaurant (i.e., the type of entity the user is searching for in that scenario) closed on this day. Entity 711 is also very close to the user's path, it is part of a very large cluster of restaurants, and it has a lot of pubs nearby. Therefore, it would be exactly what the user is searching for, despite being closed, and not being able to provide its service. A qualitative analysis of the answers given by the workers shows that the participants can be divided in four almost equally large groups:

- a first group selected entity 746 as more relevant, and specified that 711 was irrelevant because it is closed;
- a second group selected entity 746 as more relevant, but did not specify 711 as irrelevant, and did also not mention in the motivations the fact that the latter is closed;
- a third group selected entity 711 as more relevant, and specified in the motivations that they chose entity 711 despite the fact that it is closed, because of the other properties mentioned above;
- a fourth group selected entity 711 as more relevant, but did not mention in the motivation the fact that it is closed.

One could hypothesize that the workers in this last group just did not notice that the restaurant was actually closed. Moreover, the motivations given by workers in the third group specify that the entity was not relevant *per se*, but that they consider the possibilities offered in that location (the provided motivations are reported in Table C.17, Appendix C). Therefore, entities 711 and 746 have been compared in iteration X, where the description of entity 711 indicated that *"[711] is a restaurant, which is currently closed. It would take you 8 minutes to walk from your location to [711] and then to the bus station"*, without any further specifications about attributes related to cluster membership and co-location rules. The same information was also excluded from the table. In this iteration, entity 746 was judged more relevant than 711, and 711 was judged as irrelevant by the vast majority of the participants (see Table C.14, Appendix C), whereas only few workers still considered entity 711 as more relevant.

For this reason, the result of iteration X will be taken into account in the discussion of the results (see Section 7.1.3), and entity 746 will be regarded as more relevant than entity 711. I consider the latter option to be the most strongly supported by the data collected, and the other option to be less defendable. The values reported in Table 7.4 refer to the interpretation mentioned above (i.e., 746 as more relevant than 711, and 711 irrelevant).

For the sake of completeness, I also investigated four alternate ranks (see Table C.18, Appendix C), where entity 711 is considered as at last somewhat relevant (i.e., at least partially relevant, and with a proper rank value). In these four alternate ranks, I assumed that 711 would be still less relevant than 7211, which has equivalent values for the attributes related to clusters and co-location rules, but which is open (therefore 711 would also be less relevant than 724, which was judged as being as relevant as 7211). Considering each one of these alternate ranks, the correlation value between *Crowdsourced* and *ScoreGR* is lower, with respect to the correlation value reported in Table 7.4 (i.e., obtained for the rank that will be taken into account in the discussion), but the correlation remains significant, and no other assessment method gains a significant correlation with *Crowdsourced*. Therefore, the conclusions drawn in the discussion are not affected by the uncertainty that arose in iteration 1, as the same conclusions would still hold when considering one of four alternate ranks (see Table C.18, Appendix C).

7.1.3 Discussion

The findings reported above indicate that *ScoreGR* (proposed in Chapter 5 and implemented as described in Chapter 6) is an effective GR assessment method for the scenarios tested. That is, *ScoreGR* combines the criteria topicality, spatio-temporal proximity, directionality, cluster, and co-location in a single GR score that can be used to effectively rank geographic entities, in a way that resemble judgements made by humans using the same criteria. In fact, the Kendall's τ coefficients presented in Tables 7.2, 7.3, and 7.4 are significant for the correlation between the crowdsourced rank and *ScoreGR*, and for this correlation only. On the contrary, the two baselines tested do not achieve such results in any of the scenarios tested. This also shows that those simple methods are not adequate in complex scenarios.

For the first two scenarios, *ScoreGR* is able to correctly identify the most relevant geographic entity, while in the third scenario it selects the second most relevant entity as first, and vice versa (see Figures 7.1 to 7.6). *ScoreGR* also correctly identifies irrelevant entities in the first and third scenario, where those entities were spatio-temporally unavailable. On the contrary, in the first scenario, *Baseline1* and *Baseline2* select an irrelevant, spatio-temporally unavailable entity as the top-ranked one, because it is closest to the user's path. In the second scenario, three geographic entities located very close

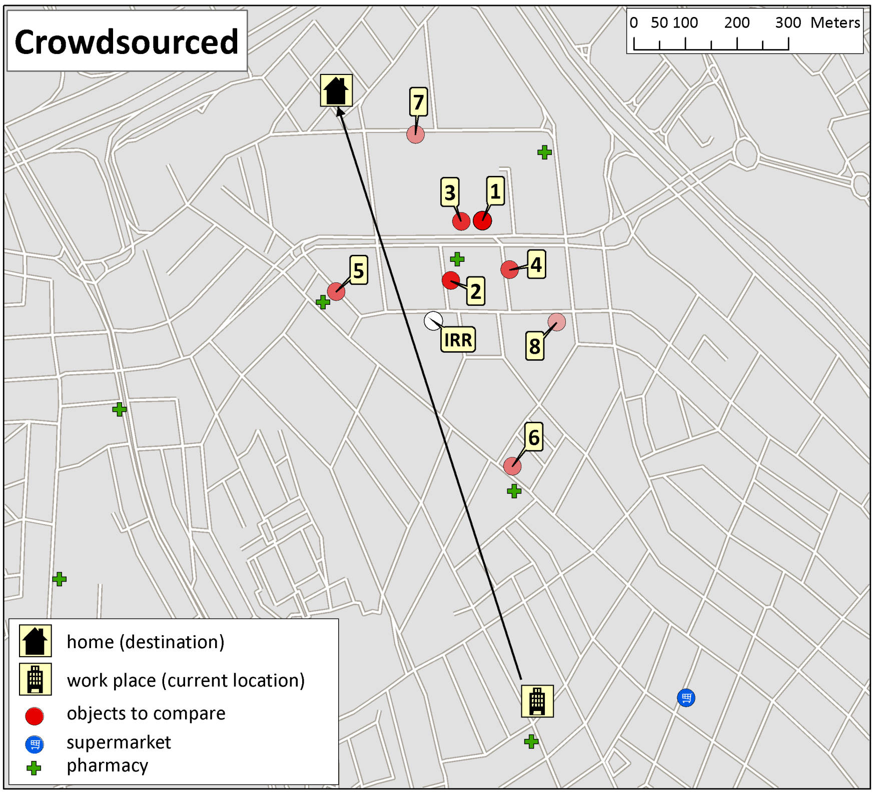


Figure 7.1: Ranking generated from the crowdsourced judgements for the entities selected with the pooling method for Scenario 1.

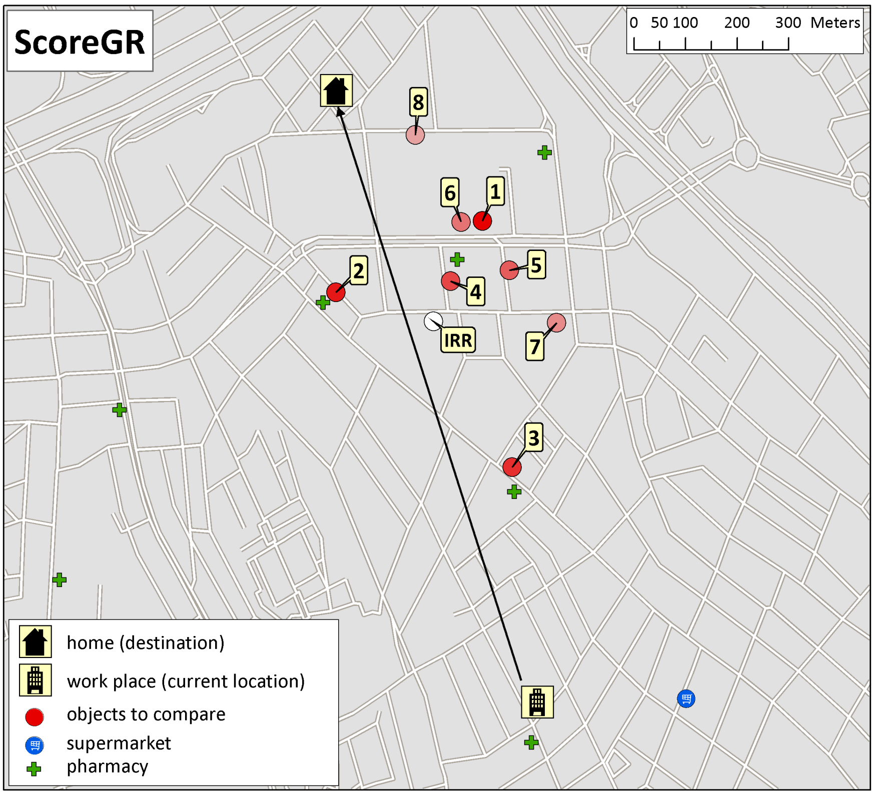


Figure 7.2: Ranking obtained by the entities selected with the pooling method for Scenario 1 using *ScoreGR*.

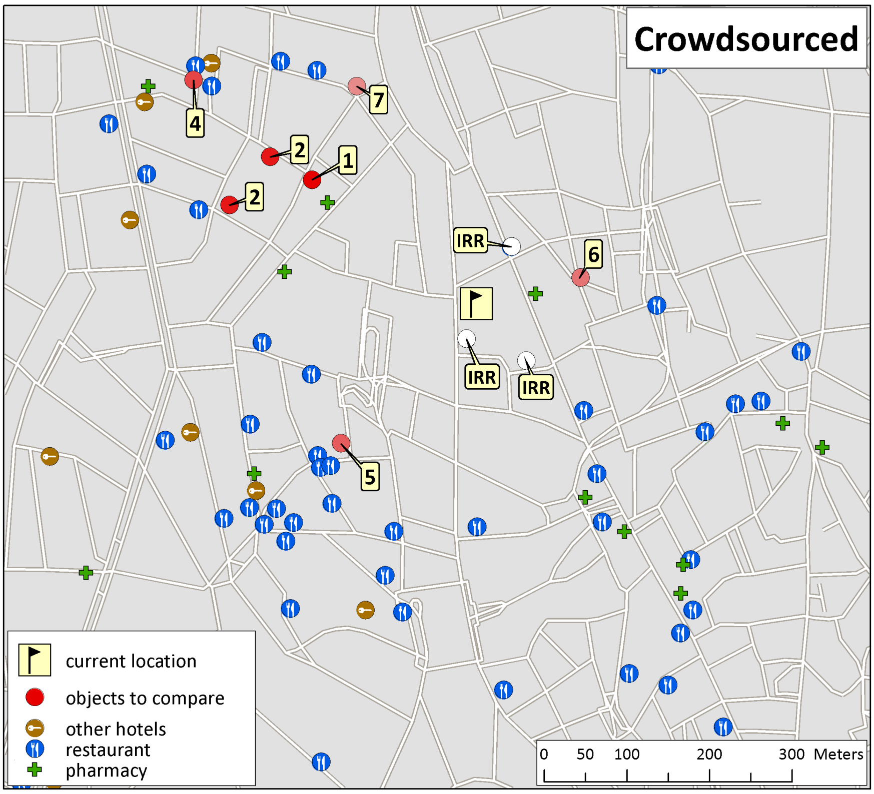


Figure 7.3: Ranking generated from the crowdsourced judgements for the entities selected with the pooling method for Scenario 2.

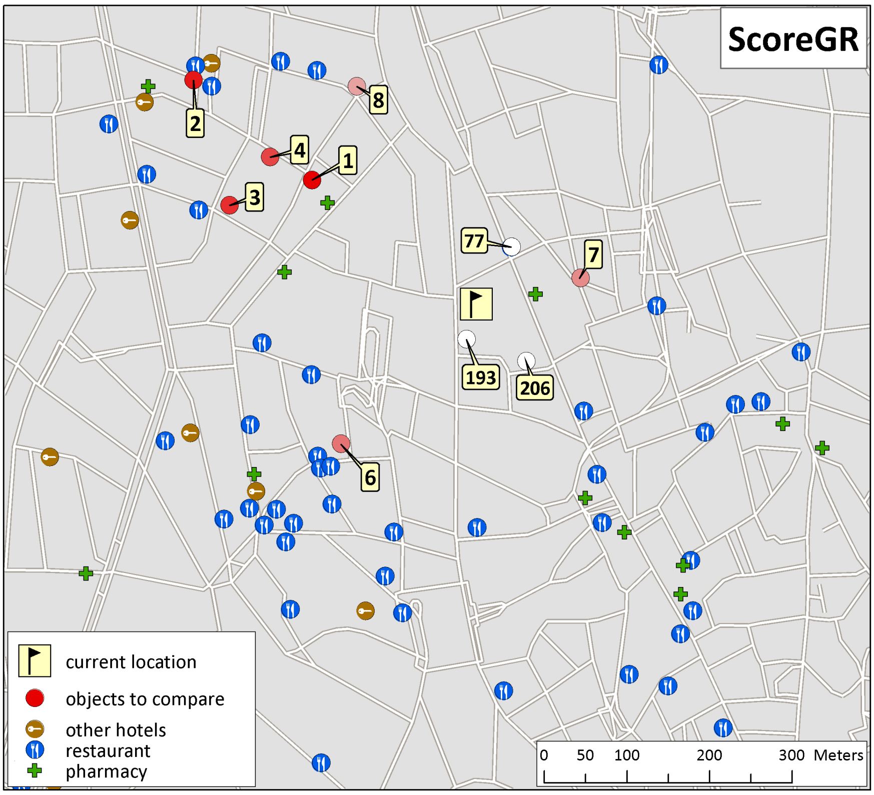


Figure 7.4: Ranking obtained by the entities selected with the pooling method for Scenario 2 using *ScoreGR*.

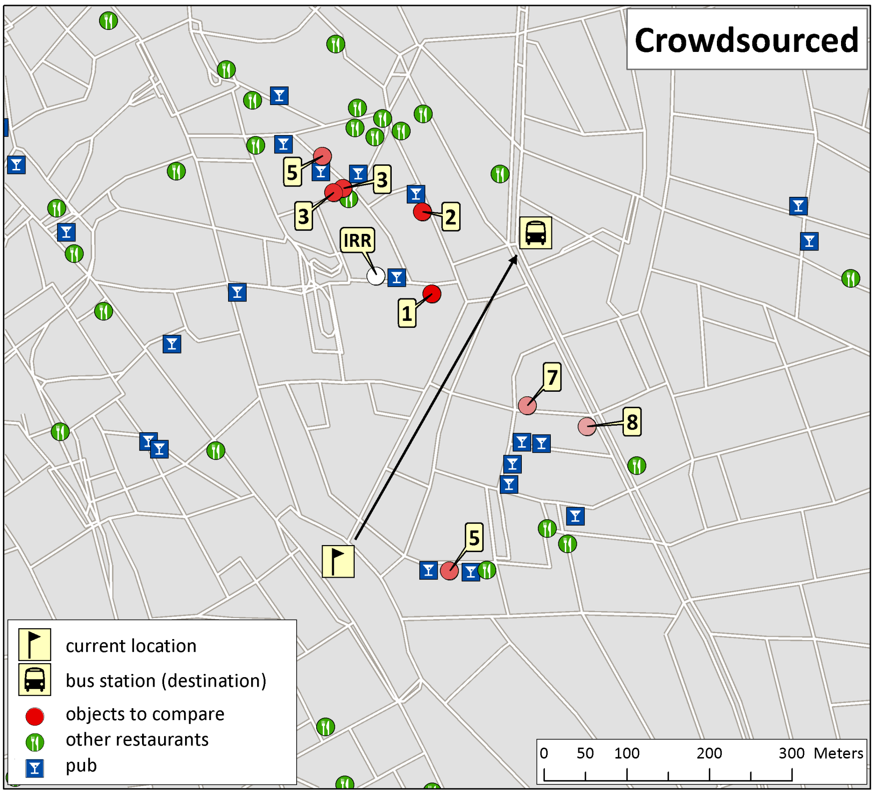


Figure 7.5: Ranking generated from the crowdsourced judgements for the entities selected with the pooling method for Scenario 3.

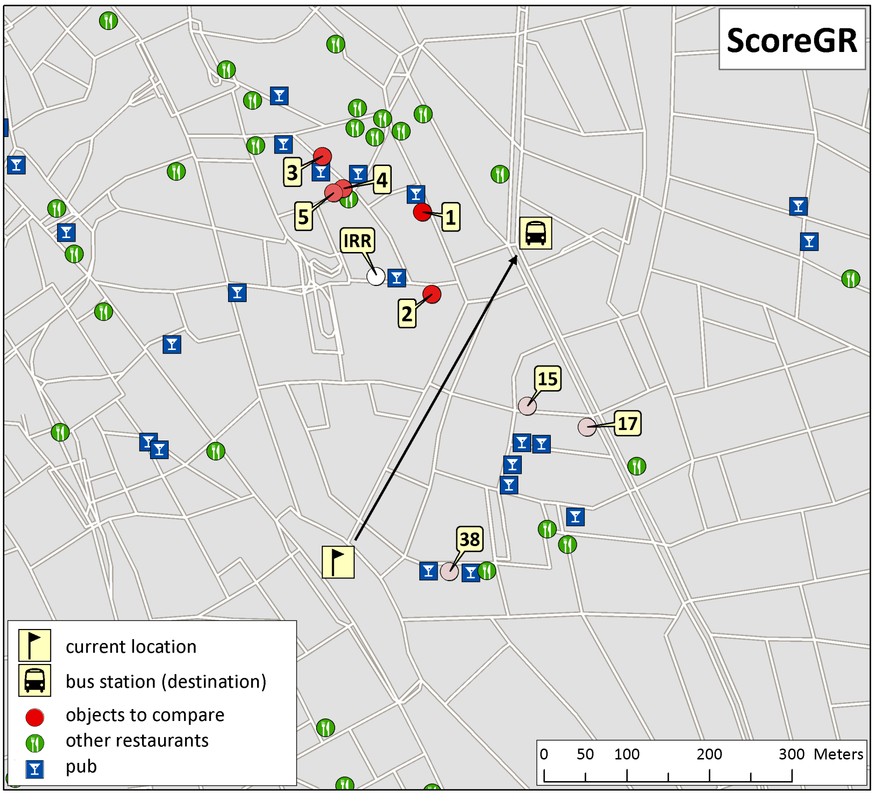


Figure 7.6: Ranking obtained by the entities selected with the pooling method for Scenario 3 using *ScoreGR*.

to the user have been classified as irrelevant by the participants, because they belong to categories not matching the user's need. *ScoreGR* does not identify these as irrelevant entities, as the measure defined in Section 6.2 still finds some semantic similarity between their categories and the user query, although the assigned ranks are very low (i.e., they are classified among the less relevant entities).

In designing *ScoreGR*, I considered as marginal the difference in the importance between the criteria cluster and co-location unveiled in Experiment I. Hence, I regarded these two criteria as equally important, when combining them to compute the geographic environment component of *ScoreGR*. As a result, the correlation coefficient between *ScoreGR* and *Crowdsource* is lower in Scenario 1 and 3, where the criterion cluster seems to be considered by participants to be more important than co-location. For instance, entity 9115 is ranked second by *ScoreGR*, but fifth by *Crowdsource* in Scenario 1, because it satisfies the co-location rule involved in the scenario well, but it is not part of a cluster. The same applies to entity 724 in Scenario 3. Considering the answers collected, participants seem to give more importance to the criterion cluster than to the criterion co-location. In all scenarios, the top-ranked entities belong to the larger cluster in the scenario, according to the crowdsourced ranks. This also resembles the results obtained in Experiment I (see Section 4.2), where the criteria cluster and co-location have been established as the sixth and tenth most important criteria, respectively. Aiming for an even better approximation of the human-based ranks, further implementations of the *ScoreGR* method will therefore require a higher importance to be assigned to the criterion cluster. Further insights into this topic may be provided by a more detailed analysis of the motivations given by workers.

Although *ScoreGR* effectively assesses GR in the three scenarios discussed above, the same does not apply to the alternate method *GRBM25*. The method was designed based on the assumption that the probability of satisfying the user's need decreases with the number of entities "closer to, or at the same distance from" the user's need with respect to each criterion (see Section 5.4). This resulted in an undesired dominance of the criteria which have a higher variability with respect to those criteria that tend to produce tied values, such as spatio-temporal proximity and clusters, respectively. For instance, the probability for a given entity to satisfy the user's need with respect to the criterion cluster is based (in part, see Section 5.3.5) on the number of other entities in the same cluster (i.e., cardinality). This value can be just one of a small set of discrete numbers from zero to the maximum value calculated for the collection, and it is the same for all the entities in a cluster. Given the definition issued in Section 5.4, this large number of entities with tied scores for a given value of cardinality (e.g., a lot of restaurants belonging to clusters of 3 restaurants) results in a equally low probability score for all those entities, creating a sort of step in the computed probability scores. On the contrary, the probability for a given entity of satisfying the user's need with respect to the criterion spatio-temporal proximity is based on the time a user is able to spend at the entity's location. This value is different for each entity depending on its distance and temporal availability. This results in a more gradual decreasing of the probability scores. Therefore, the probability scores computed for the criteria topicality, spatio-temporal proximity, and directionality have a dominant influence on the final score produced by *GRBM25* with respect to the probability scores computed for the criteria cluster and co-location. This has been confirmed by computing an additional baseline method, which filters the geographic entities by category and ranks them based on their spatio-temporal proximity score (see Section 5.3.2). In fact, there is no correlation between the crowdsourced ranks and this additional third baseline, but the latter shows a highly significant correlation ($p < .01$) with *GRBM25* for all the three scenarios (with Kendall's correlation coefficients $\tau > .90$ for Senarios 2 and 3). This result also confirms that the criteria cluster and co-location criteria have to be treated separately, as they cannot be captured using simplistic spatial criteria.

The results obtained also highlight the risks entailed in combining thematic and spatial scores. In Scenario 2, geographic entities very close to the user (i.e., with high spatial score) are ranked as very important by *Baseline2*, even if they are not what the user is searching for (i.e., low thematic score). *Baseline1* is obviously not affected by this particular problem, as the rank is solely based on spatial distance. However, *Baseline1* is still not effective in the scenarios tested, and it suffers from the issues related to a filter-based method (e.g., motel and hostel would be considered as completely irrelevant, when searching for a hotel). Even considering just one thematic criterion and one spatial criterion, a combination method based on continuous logic may be a more effective solution. For instance, methods such as the Continuous Preference Logic (CPL) model (Dujmovic, 1975, 2007) used in *ScoreGR* allow fuzzy conjunctive combinations and conjunctive partial absorptions, and avoid the issue related to other conjunctive combination methods (as discussed Section 5.5). Therefore, the CPL may be more adequate than the geometric combination method used in *Baseline2*, which is a disjunctive combination.

Future studies could focus on evaluating the effectiveness of the methods considered in scenarios where the criteria cluster or co-location are not as important, in order to test for the robustness of the proposed methods. As already mentioned in this chapter, *ScoreGR* was designed so that those criteria have a neutral influence on the GR score of entities that belong to categories not involved in cluster and co-location rules. Therefore, future evaluations could focus on scenarios involving categories usually involved in cluster and co-location rules, which should however not be important in the specific selected situation. For instance, the discussed methods can be tested in a scenario depicting a user searching for a hotel, in an area where there are hotels, but there are no hotel clusters, and none of the hotels satisfies any of the co-location rules.

7.2 Summary

This chapter reported on Experiment III, where I investigated the effectiveness of two different GR assessment methods and two baselines, following the approach used by Urbano et al. (2010), based on crowdsourcing. A pooling system was used to select the geographic entities to be used in the three scenarios prepared for the "user-centred" evaluation procedure. The crowdsourced data were used to produce a "ground truth" rank, which was compared with the four methods used in the pooling phase. The results support the effectiveness of the main GR assessment method *ScoreGR* (proposed in Sections 5.3 and 5.5, and implemented as described in Chapter 6), and the inadequacy of the other three tested methods.

The next chapter will offer an overall discussion of the outcome of this dissertation. In particular, I will answer the research questions issued in Chapter 1, discussing in detail the overall implications Experiments I, II (see Chapter 4), and III for the development of the concept, criteria, and assessment methods of GR.

Chapter 8

Discussion

In the following sections, I first provide a general answer to the main research question issued in Chapter 1, and detailed answers to the four research questions issued in Section 1.2. Thereafter, I discuss the outcome of this dissertation in the broader context of GIScience and IR in Section 8.2, analysing the implications of the reported findings in the scope of the current understanding of relevance in mobile information services. Finally, the limitations of the empirical studies presented in this dissertation are described in Section 8.3.

8.1 Answering the research questions

The main research question pursued in this dissertation was posed in Chapter 1:

- *Which information and criteria are needed, and how do they have to be combined in order to assess a set of numerical values that estimate the GR of a given geographic entity with respect to a given context of use?*

Given the results obtained in Experiments I, II (see Chapter 4) and the outcomes of Experiment III (see Chapter 7), I argue that GR is a complex and novel concept, which expresses the multi-faceted relationship between a user's geographic information need, and geographic entities in the surrounding environment. The two main criteria defining the strength of the GR relationship are the spatio-temporal accessibility of an entity with respect to a user's mobility (encapsulated in the criterion spatio-temporal proximity), and the topicality of an entity's affordances with respect to a user's activity and information need (encapsulated in the criterion topicality). Furthermore, the strength of the GR relationship is strongly affected by the geographic context in which an entity is placed, including first- and second-order effects, such as spatial clusters and co-location rules. Therefore, GR is different from the concept of relevance commonly used in IR. Finally, the empirical evaluation presented in Chapter 7 shows that GR cannot be adequately calculated as a simplistic combination of category filtering and distance-based ranking, whereas the main method *ScoreGR* presented in Chapters 5 and 6 has proved to be effective in assessing GR in the scenarios considered.

8.1.1 Modelling

- **RQ1**: *Which information is needed to model the user context, the geographic entities, and the surrounding environment in order to assess GR?*

In order to assess GR with respect to the user context, the geographic entities, and the surrounding environment, it is necessary to account for information about the user's geographic information need and activity, along with the user's personal preferences and mobility. Also, it is necessary to account for information about the affordances related to the geographic entities, their attributes, location and temporal availability. It is furthermore necessary to consider the social and computational context in which the user is acting, along with the geographical context in which the entities are placed, that is the geographic phenomena they are part of. Finally, it is important to incorporate information about the context in which a user's activities take place, including the prior, subsequent, and co-occurring activities, the user's overall objective, and planned mobility.

The information needed to model the user context, the geographic entities, and the surrounding environment in order to assess GR is summarised in the conceptual model presented in Figure 3.2 (see Chapter 3).

- **RQ2**: *Which criteria can be used to assess the relevance of geographic entities?*

The outcomes of Experiments I and II reported in Chapter 4 show that criteria commonly used in IR, such as topicality, accuracy, and currency, are important criteria of GR. However, those criteria are clearly not sufficient for understanding GR, as the facet concerned with the user's mobility acquires a crucial role in mobile information services. The answers provided by participants also highlight the inadequacy of the concept of relevance commonly implemented in current mobile information services. Taking into account spatial distance alone is not sufficient, but a combined conception of space and time is fundamental in assessing GR. The relevance of a geographic entity is not simply related to its distance from the user, but eventually to the time the user will be able to spend at that location with respect to the time needed to perform a planned activity. This is also supported by the results of Experiment III presented in Chapter 7, where entities close to the user, but temporally unavailable are considered as irrelevant. At the same time, other factors related to the surrounding environment can interfere with this notion of spatio-temporal relevance, as discussed below.

Accounting for the context in which a user is interacting with an information system has been one of the main topics in computer science and related fields in the last twenty years (Schilit et al., 1994). This dissertation supports the idea that not only this notion of context matters, but also the context in which a geographic entity is placed is a crucial factor in GR. The influence exerted by the criteria cluster and co-location on the ranking of geographic entities has been proved by the results obtained in Experiment

I and II (see Chapter 4) and confirmed by the results obtained in Experiment III (see Chapter 7). The criteria cluster and co-location are not independent from other geographic criteria (e.g., entities in the same cluster will be at similar distances from the user, probably have similar temporal availability, and therefore similar spatio-temporal relevance), but the geographic context of the entities is not entailed by simplistic spatial criteria. The criteria cluster and co-location criteria have to be treated separately. In fact, the baseline methods tested in Experiment III were not able to resemble the human-made judgements, which were considering the criteria cluster and co-location. The baseline methods perform better in Scenario 2, where the temporal component of spatio-temporal proximity is not important, than in Scenarios 1 and 3, where the temporal component is essential. In Scenario 2 all entities are equally temporally available, and the user's location coincides with the user's destination, thus spatio-temporal proximity is reduced to spatial distance, and directionality is excluded. Hence, in Scenario 2 the mobility component is reduced to the same spatial distance criterion taken into account by the baseline methods. However, even in Scenario 2 the ranks produced by the baselines methods do not resemble the ranks obtained from the collected human-made judgements.

In some cases, the relevance of the environment in which a geographic entity is placed seems to overcome even the importance of the criterion spatio-temporal proximity. Some of the participants in Experiment III disagreed on the irrelevance of an unavailable entity, although they were not the majority. This kind of judgement may seem odd at first, but the motivations issued by those participants actually provide an insight into the human geographic decision making process. Some participants agree that a temporally unavailable entity may still be relevant, and even more relevant than another entity, if the surrounding environment is relevant. The motivations given refer to the opportunities that the participants would have, once they found themselves at that location. For example, a closed restaurant could be still relevant if there are other open restaurants, and pubs nearby (see Scenario 3, Experiment III, Chapter 7). At the same time, a closed supermarket with no other supermarket nearby is indeed irrelevant (see Scenario 1, Experiment III, Chapter 7). Nonetheless, it can be doubted that users would consider a service effective if it suggests closed restaurants as good places to go for dinner.

A further insight into the human geographic decision making process is offered by another interesting group of answers provided in Scenario 1 of Experiment III (see Chapter 7). In that scenario, the user was searching for a supermarket. It seems reasonable to think that a supermarket where the user would have more time for shopping would be more relevant. For example, assuming that the user would need at least 20 minutes for shopping, and that she has to be at the destination in about one hour, a supermarket where the user would have about 50 minutes for shopping (see entity A in Figure 8.1) should be more relevant than (or at least as relevant as) a supermarket where she would have only about 30 minutes time for shopping (see entity B in Figure 8.1). Nevertheless, in the specific case the two supermarkets are rather close to each other, and the second is located just between the first and the user's destination. It would be possible for the

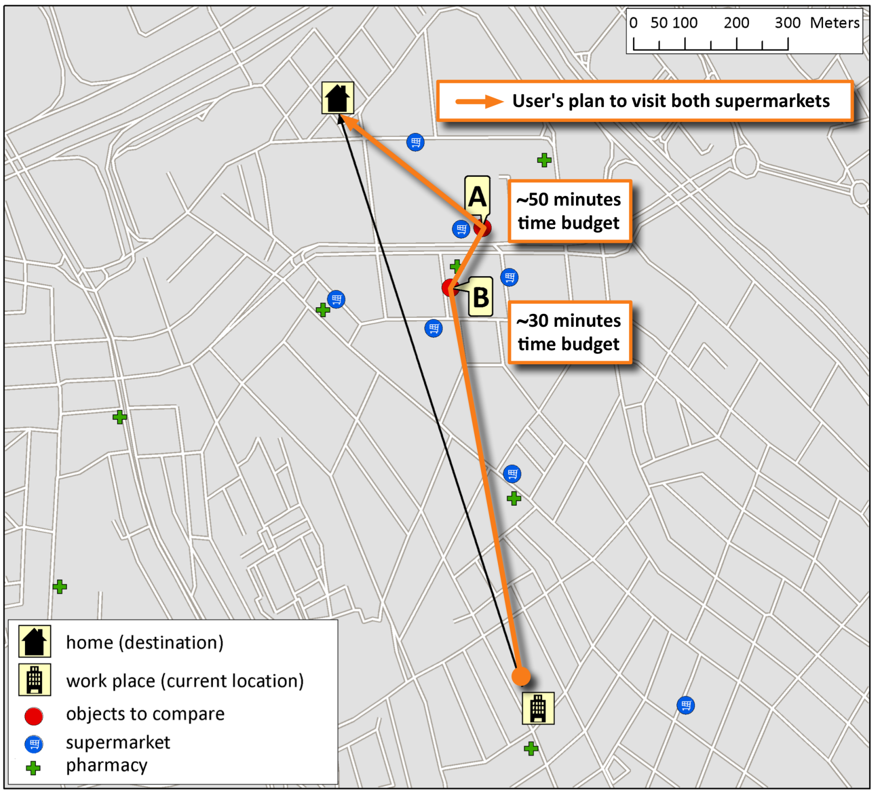


Figure 8.1: Illustration of the example related to the criterion planning.

user to go to the supermarket with 30 minutes time budget, and pursue the shopping activity. The user could then continue towards the destination, and still have the possibility to stop by the other supermarket (see the orange path in Figure 8.1), with 20 minutes left in the time budget, in case any item is not available in the first supermarket. Therefore, the second entity (see entity B in Figure 8.1) would be a better choice, in this perspective, as first supermarket to go to for shopping. It would be interesting to consider a hypothetic situation in which the two time budgets are switched, and investigate whether a user would accept to walk back to the second entity (from A to B in Figure 8.1).

Based on the two insights described above, I propose a new criterion *planning*, which refers to the extent to which reaching the entity can favour the user's plans in accomplishing an activity. This new criterion of GR is included in Table 8.1, which updates the list of criteria of GR presented in Chapter 4 (see Table 4.1) and De Sabbata and Reichenbacher (2012). The two insights described above also suggest that in modelling the relevance in mobile information systems, the geographic environment in which the entities under assessment are placed is as important as the context in which the user is seeking information to satisfy her needs. An important next step in this line of research will be to consider the role of the context in which the user's activity is placed, in terms of prior, subsequent, and co-occurring activities, which may have an important influence

Table 8.1: Criteria of GR.

Properties	Geography	Information	Presentation
topicality	spatial proximity	specificity	accessibility
appropriateness	temporal proximity	availability	clarity
coverage	spatio-temporal proximity	accuracy	tangibility
novelty	directionality	currency	dynamism
popularity	visibility	reliability	presentation quality
	anchor-point proximity	verification	
	hierarchy	affectiveness	
	cluster	curiosity	
	co-location	familiarity	
	association rules	variety	
	planning		

on the perceived relevance of the geographic entities under assessment. As geographic entities *"do not exist as independent entities, but rather they exist within a specific geographic context"* (De Sabbata and Reichenbacher, 2012, p. 1500), human activities do not just happen, but rather they are part of a continuous succession of interrelated events.

It is worth noting that the examples provided above are indeed peculiar situations, and the type of answer discussed above was provided only by a relatively small number of participants. Moreover, taking into account the criterion *planning* would mean to consider each other entity as second possible stop, squaring the number of possible options to compute. If more than two stopovers are be considered, the problem becomes exponential in terms of numbers of possible solutions. In fact, such a problem resembles a multi-object recommendation (see e.g., Schlieder, 2007) or a tour planning problem (see e.g., Seifert, 2007), which is equivalent to solve a finite domain constraint satisfaction problem (known to be NP-hard, see e.g., Russell et al., 1995). A future perspective on the planning criterion could be to integrate a planning algorithm such as the one proposed by Abdalla and Frank (2012). Each geographic entity under assessment could be considered a starting point for a subsequent activity, or an ending point of a prior activity, and be judged based on the quality of the solution computed by the planning algorithm (e.g., in terms of time needed or distance travelled). Considering both a prior and a subsequent plan, a geographic entity could also be judged as the middle point of a larger activity.

8.1.2 Assessment

- **RQ3**: *How can we use information about user context, geographic entities, and the surrounding environment to compute a numerical value for each criterion of GR?*

The results reported in Chapter 7 confirm that the scores proposed in Section 5.2 are an adequate means to compute a numerical value for the criteria topicality, spatio-temporal proximity, directionality, cluster, and co-location, in the scope of assessing

GR. However, there are two noteworthy points related to the computation of two of the criteria mentioned above, that emerge from Experiment III.

The first issue highlights the importance of an adequate estimation of topicality. If the only description provided for the geographic entity under assessment is the name of the category it belongs to, it is not straightforward to provide a non-binary score for topicality (i.e., not just a filter). Although filtering entities by category can be an efficient solution, and an effective one as well in many cases, in other cases the user may be interested in "similar" entities – a nearby motel versus a distant hotel. The problem is how to adequately define and implement such similarities. The solution proposed in Section 6.2 is reasonable and produces meaningful scores for topicality, but it is strongly dependent on the underlying dataset. It can also produce scarce results, where pubs are almost as similar to hotels as hostels are to hotels, as illustrated in Figure 6.1. The task becomes even more difficult, if the user specifies her objective (i.e., a planned activity) instead of a category name (e.g., "dinner" instead of "restaurant"). In the scope of the project described here, there has been no opportunity to test the implemented prototype in this specific direction, so no conclusion can be drawn about this point. In general, a possible solution to such problems can be to actually search for the objective behind the user query, and match it with the affordances related to the entities. However, such a solution would require a sophisticated software with a very good understanding of human common-sense (Singh et al., 2002; Liu and Singh, 2004; Gunning et al., 2010), especially concerning the affordances associated with different types of geographic entities (Alves and Pereira, 2012; Alazzawi et al., 2012).

A second issue is related to the assessment of spatio-temporal proximity. It has been shown (see Section 6.3 and Boscoe et al., 2012) that Euclidean distances are a good estimation of the distance in the route network. Therefore, this simple measure can be effectively used to efficiently compute an approximate value for spatio-temporal accessibility. However, this model holds only as far as a uniform speed of movement is assumed. Even in the case of a tourist walking through a city centre, it would be appropriate to take into account the use of a public transport network. A good estimation of travel time and accessibility should account for pedestrian, bicycle, and route networks including parking areas, public transport facilities, and traffic information, and allow for multi-modal transportation. Needless to say that performing traffic-aware multimodal routing for each of the geographic entities considered in a GR assessment would imply a considerably high computational cost, and would still be just an estimation of spatio-temporal accessibility, not matching reality. It is therefore important to investigate an appropriate balance between the accuracy of the estimation, and the computational cost, in order to provide a reliable service, and short response time.

- **RQ4**: *How can we combine the values representing single criteria of GR in an adequate way?*

Based on the results of Experiment III, I argue that the score-based method *ScoreGR* proposed in Chapter 5 performs an effective assessment of GR, in the scope of the five selected criteria. The method is able to produce ranks similar to those derived from human-made judgements, achieving significant correlations with coefficients in the range of .50 to .85[1], depending on the scenario. The correlation coefficients are lower for those scenarios in which the criterion co-location is less important than the criterion cluster (according to the ranks derived from human-made judgements), as *ScoreGR* has been designed to assign equal weight to these two criteria. This difference suggests that the criterion cluster plays a more important role than the criterion co-location in assessing GR, but also suggests that the importance of the criteria can be different in different scenarios. At the same time, the results obtained in Experiment III suggest that both the criteria cluster and co-location play a more important role in the computation of the final score, with respect to the importance assigned to those criteria in the combination method implemented in *ScoreGR*. In fact, the entities which better fulfil those criteria are ranked higher in the crowdsourced rank than in the rank obtained using *ScoreGR*.

The results reported in Chapter 7 confirm the issues raised in Section 5.5, concerning the method used to combine the scores computed for different criteria of GR. The drawbacks of simple conjunctive and disjunctive methods have been described and confirmed. Given the observed poor functioning of such methods when applied to only two criteria (topicality and spatial distance, which are also supposedly equally important in assessing GR) it can be argued that such methods cannot be applied to a larger number of criteria, each one with a different role and importance. The Continuous Preference Logic (CPL) model (Dujmovic, 1975, 2007) seems instead to offer the expressiveness required to adequately model the combination of the criteria. The use of partially conjunctive logic operators capable of handling both mandatory and desired criteria resulted in an effective combination schema (see Figure 5.6), which also allows for an easier conceptual framework with respect to a list of weights to apply in a linear combination. Nevertheless, the selection of the strength of the "and-ness" could also be as arbitrary as the selection of the weight of a linear combination. A similar approach of prioritised aggregation of multiple criteria by means of a prioritised "and" operator has been recently proposed by da Costa Pereira et al. (2012).

8.2 General discussion and scientific contribution

The outcomes of this dissertation not only provide empirical evidence supporting the critiques of simplistic approaches to relevance, which are still common in current commercial LBS applications and mobile information services, for instance as formulated by Reichenbacher (2004) and Raper et al. (2007b), but also offer effective solutions based on the concept of GR (Raper, 2007; Reichenbacher and De Sabbata, 2011). In this dissertation, I address the "manifestation" of GR, that is the fourth domain of GR proposed by Raper (2007), as an egocentric place-oriented concept of relevance. The perspective on relevance is egocentric, as it accounts for a user's mobility and spatio-temporal ac-

[1]These values refer to the Kendall's τ correlation coefficient (Kendall, 1938).

cessibility, and place-oriented, as it accounts for a user's activity and the affordances related to geographic entities. At the same time, I employ a pragmatic definition of GR, similar to the concept of relevance defined by Saracevic et al. (1988), in order to develop a computational method to assess GR as the logical consequence of the data available within a computer system, as discussed in Chapter 3.

On the one hand, the findings reported in this dissertation confirm the main assertion by Raper (2007, p. 837), that *'situational relevance concepts as currently articulated do not deal sufficiently with concepts of mobility and geography, and that these concepts are essential to the understanding of mobile information seeking'.* Current commercial LBS applications focus on spatial proximity, as the sole geographic aspect of relevance, as did most of the earliest approaches proposed in literature (for a survey, see e.g., Raper et al., 2007a). The empirical studies presented in this dissertation show that simplistic approaches, which use distance-based filtering or ordering, are not adequate for the assessment of relevance in mobile information services. These simplistic methods are not applicable, even to simple user requests, such as "show me the closest supermarket". I provide empirical confirmation that spatio-temporal proximity is a fundamental criterion of GR. Mobile information services need to account for at least a basic understanding of time geography (Hägerstrand, 1970; Pred, 1977; Miller and Bridwell, 2009; Kuijpers et al., 2011), as first suggested by Mountain (2005), and discussed by Raubal and Panov (2009) and Crease (2012). Although not explicitly specified, in a typical user request, such as "show me the closest supermarket", there is an implicit requirement for the user to be able to perform her activity, and thus for the entity to be spatio-temporally accessible, considering both the user's mobility, her time schedule, the location of the entity, and its temporal availability.

The effectiveness of LBS applications might also be affected by issues related to the common simplistic methods used to combine different scores of relevance, such as the methods proposed by Zipf (2003) and Reichenbacher (2004). As illustrated in Chapter 5, one of the most common combination method, arithmetic summation has been excluded from the design of the presented GR assessment method *ScoreGR*, because of its compensatory nature. These issues have been confirmed by the results obtained in the empirical study presented in Chapter 7. Geometric combination has also been found to be inadequate to combine topicality and spatial distance in a mobile information seeking scenario, although it has been effectively applied in GIR (see Purves et al., 2007). I argue that mobile information services can profit substantially from exploiting models based on continuous preference logic (Dujmovic, 1975, 2007), for combining the different criteria used in the relevance assessment.

On the other hand, the findings reported in this dissertation contradict the idea suggested by Raper (2007, p. 842–843), that *'whatever is in the [accessibility envelope or surroundings of the user] is also relevant topically, simply because it is in the [accessibility envelope or surroundings of the user]'.* The relevance judgements collected in Experiment III (see Chapter 7) clearly show that a geographic entity, which is topically non-relevant, is judged as non-relevant, even if it is very close to a user's position.

Moreover, as *'relevance is not necessarily the same as topicality'* and *'a document on a different topic might, for one reason or another, satisfy the user's information need'* (Bookstein, 1979, p. 270), GR is not necessarily the same as spatial-temporal proximity, and a geographic entity might, for one reason or another, satisfy the user's geographic information need better than another geographic entity closer to the user's position, or available for a longer period of time. The collected evidence shows that geographic first- and second-order effects (O'Sullivan and Unwin, 2003) have a substantial influence on the relevance of geographic entities, as first suggested by Reichenbacher (2004). The geographic environment surrounding the entities under assessment is part of the context component of relevance (Mizzaro, 1998; Coppola et al., 2004), and it needs to be included when defining concepts of relevance addressing real/physical world entities, such as the concept of w-relevance suggested by Coppola et al. (2004), and the concepts of relevance employed in LBS (e.g., Hauthal and Burghardt, 2012; Huang and Gartner, 2012; Pombinho et al., 2012), spatial Web objects IR services (e.g., Cao et al., 2010; Venetis et al., 2011), and recommender systems (e.g., Saiph Savage et al., 2012; Barranco et al., 2012). Mobile information services can substantially improve their effectiveness by taking into account criteria related to the geographic environment. The criteria cluster and co-location have been found to be of particular importance, in assessing GR. The outcomes of this dissertation confirm that the relevance of a single entity increases, if there are several entities of the same kind in the neighborhood (Reichenbacher, 2005a), or if it satisfies a common co-location rule (De Sabbata, 2010). These criteria are the '*the differences in situational contexts and research task requirements*' (Barry and Schamber, 1998, p. 234) that differentiate GR from the concept of relevance commonly applied in classic document-based IR. This finding also supports the validity of the notion of prestige, proposed by Cao et al. (2010) as a measure of centrality of a spatial Web object within a cluster of similar and correlated objects. The notion of prestige could be integrated in GR as a further facet of the criterion cluster.

Moreover, as discussed above in Section 8.1.1, I advocate the importance of taking into account, not only the context in which a user is while seeking geographic information, and the geographic environment in which entities exist, but also the context in which a user's activities take place, for computing the relevance of geographic entities with respect to a user's information need. Based on a preliminary qualitative analysis of the motivations provided by the participants in the empirical study presented in Chapter 7, I argue that a mobile information service should account not only for a single activity, but also for the context in which an activity happens. The relation between an activity, which a user is seeking information for, and other previous, co-occurring, or subsequent activities should be taken into account. The relevance of an entity could then be calculated considering the consequences of a user performing an activity at that place with respect to a user's overall plan. Personal information management (Abdalla and Frank, 2012), task planning (Seifert, 2007), and multi-object recommendation (Schlieder, 2007) could provide a framework for the future development of the new criterion *planning*. The latter is included in Table 8.1, which updates the criteria table first proposed in De Sabbata (2010) and revised in De Sabbata and Reichenbacher (2012). At the same time, further studies are necessary to understand the influence of GR on those three fields of research, and on the users' mobile decision making processes in general. These studies should be accompanied by further research on the applicability of activity theory (Kaptelinin and Nardi, 1997, 2006) in the scope of mobile information services (Reichenbacher, 2005b; Dransch, 2005; Huang and Gartner, 2009).

Finally, I advocate for a better understanding of the affordances of geographic entities (Alves et al., 2009; Alves and Pereira, 2012; Alazzawi et al., 2012), which is crucial in defining the topicality of a geographic entity with respect to a user's information need, as discussed in Chapter 6. Moreover, information about the temporal aspect of the affordances related to an entity (Alazzawi et al., 2012) is necessary, as soon as the criteria spatio-temporal proximity and planning are taken into account in assessing GR.

8.3 Limitations

As mentioned in Chapter 3, due to the limited temporal extent of the project, only five of the most important criteria of GR (according to the results reported in Chapter 4) presented in Table 8.1, among the 31 listed, have been studied in detail. This selection focused on the novel part of the criteria list, as it concerns the criteria I presented in De Sabbata and Reichenbacher (2012), and it relates to the geographic environment surrounding the entities under assessment, which is of most interest for the field of GIScience. At the same time, it is not possible to foresee the influence of important criteria, such as accuracy and currency (see Chapter 4), on the GR assessment method, as they have been excluded from the selection. For the same reason, it is also difficult to speculate on the role that components related to a user's preferences and social network information can have on GR. These components have been included in the conceptual definition of GR, but have been excluded from the empirical studies presented in this dissertation. These two are the pivotal components of recommender systems, commonly and effectively used in a large number of commercial systems. Further studies are needed to investigate the interaction between personal preferences and geographic criteria in the scope of GR, and it is possible that the first could be more important than the second.

It must be noted that the empirical studies presented in this dissertation were conducted in the form of Web-based questionnaires; the third experiment using the crowdsourcing platform CrowdFlower[2]. Therefore all the results collected refer to hypothetical situations. This research did not include any field experiment for the following reasons. First, a field experiment does not offer a controlled environment appropriate to the stage of development of the GR assessment method under evaluation. In fact, personal preferences (e.g., fish&chips over pizza) and habits (e.g., familiar over little known areas of a city) resulting from prior knowledge of the geographic environment can have a considerable influence on a user's perception of relevance, as discussed in Chapters 3. Criteria

[2] `https://crowdflower.com`, last accessed September 2012

related to personal preferences and habits have not been included in the GR assessment method proposed in Chapter 5, but they would still influence the participants' perception of relevance in the case of a field study, and create significant biases. These biases would complicate the study of general criteria of GR, which was the aim of the presented empirical studies. Further studies are needed to investigate the relation between the studied criteria, and the criteria related to personal preferences and habits. Second, the development of a fully functional mobile application for evaluating the assessment methods proposed in Chapter 5 would have posed software engineering issues. The most prominent would probably have been the response time of the developed prototype, which is not comparable to current mobile applications, and search engines standards. This would in turn strongly affect and bias participants' opinion about the effectiveness of the service. Furthermore, the scenarios used in the presented experiments were set in an urban environment. The possibility that different settings might require different criteria has to be taken into account. For instance, the criterion visibility was well rated by the participants of the study conducted by Mountain and Macfarlane (2007) in the Swiss National Park, whereas the same criterion has been rated as one of the least important by the participant of the first experiment presented in Chapter 4.

Finally, the data collected in Experiment III (i.e., concerning which one of two proposed entities is more relevant, and why, see Chapter 7) have been mostly quantitatively analysed. Little attention has been paid so far to a qualitative analysis of the motivations provided by each participant for each judged comparison. In the scope of this thesis they have only been used in order to unravel unclear or contradictory judgements (see Section 7.1.2. Instead, once properly formatted and deployed to the public, this data can be a valuable contribution to projects such as MIRA (Allan et al., 2012), which aims at developing community-wide benchmark tasks, test collections, and evaluation methodologies in the field of mobile IR. In fact, this data can be used to create a benchmark for testing other GR assessment methods and LBS applications. Moreover, the dataset contains more than two thousand short textual motivations that can be an excellent starting point to gain a deeper understanding of GR and mobile geographic decision making processes.

Chapter 9

Conclusion

9.1 Achievements

The basic hypothesis of this research is that Geographic Relevance (GR) is a novel concept, which differs from the conceptualisation of relevance commonly used in IR, and therefore a novel computational method is needed to assess GR. This dissertation provides an in depth study of GR, from conceptual development to prototype implementation, and evaluation of the proposed GR assessment method. The derivation of the concept from previous conceptualisations of relevance in the fields of IR and GIScience has been presented. A survey of the criteria necessary to assess GR has been conducted, and novel, important criteria have been uncovered, which have their roots in geographic data mining and geographic information analysis. A novel GR assessment method was proposed, and prototype software has been developed accordingly. It has been shown that this method is capable of effectively estimating GR, whereas baseline methods resembling simple approaches widely used in common LBS applications nowadays were not successful, confirming with empirical evidence previous research in the fields of LBS (Schmidt et al., 1999; Reichenbacher, 2005b; Jiang and Yao, 2006; Raper et al., 2007b).

The aim of this research was to define an assessment method for numerically estimate GR (see Chapter 1). For this purpose, it is fundamental to focus on the relation between relevance and geography, where the latter refers mainly to the user's mobility in space-time, and the geographic environment surrounding the entity under assessment. These are important facets of the context in which the relevance relationship between a user and a geographic entity manifests itself, but they are not the only ones. A review of the different fields which treat the relation between relevance and geography is presented in Chapter 2, along with the main concepts discussed in this dissertation. The same chapter also covers geographic data mining and geographic information analysis, as well as system evaluation, including crowdsourcing, which was applied in the presented research to estimate GR and evaluate the proposed approach.

The conceptual development of GR in the scope of previous and similar notions of relevance in IR and GIScience is presented in Chapter 3. First the elements involved in the GR relationship between a user, a geographic entity, and the surrounding environment were identified. Thereafter, the literature was reviewed for criteria of relevance which could have been useful in analysing the relationships between those elements. Moreover, new criteria were suggested which are related to first- and second-order effects in the geographic environment surrounding the entities (i.e., the criteria hierarchy, cluster, co-location, and association rules), and a user's understanding of space and places (i.e., anchor-point proximity). The relative importance of all these criteria in GR have been investigated in two empirical studies presented in Chapter 4. Given the results obtained, the hypothesis of equivalence between GR and the concept of relevance commonly employed in IR has been rejected, because of the clear differences in the importance of the criteria in these two different conceptualisations. In particular, the criteria hierarchy, cluster, and co-location were found to be of primary importance in GR, although being novel with respect to the previous conceptualisations of relevance in IR and GIScience.

Common methods for relevance assessment are not sufficient to assess GR, as they do not encompass the new criteria of GR discussed in this dissertation (i.e., the criteria hierarchy, cluster, co-location, association rules, and anchor-point proximity). Therefore, in Chapter 5, I propose a new method to assess GR, which is referred to as *ScoreGR*. This includes the criterion topicality, which is the fundamental criterion of relevance, along with the criteria spatio-temporal proximity and directionality, which entail the user's mobility, and the criteria cluster and co-location, which offer information about the geographic environment as a distinctive facet of GR (see Section 5.3). The method presented uses continuous preference logic (Dujmovic, 1975, 2007) to obtain a fuzzy conjunctive combination of the calculated scores (see Section 5.5). *ScoreGR* has been implemented in a prototype software, along with an alternative GR assessment method referred to as *GRBM25* (see Sections 5.4 and 5.5), as described in Chapter 6. In order to investigate the effectiveness of the proposed methods, I set up an empirical study, using the procedure followed by Urbano et al. (2010) to generate "ground truth" ranks from crowdsourced relevance judgements for three different scenarios (see Chapter 7). The results of the empirical study show that *ScoreGR* is an effective method to assess GR. The same results show that the two baselines taken into account (which resemble simple approaches commonly used in LBS) and the alternative method *GRBM25* do not effectively estimate GR in the considered scenarios.

Hence, in order to assess GR it is necessary to model the information included in the conceptual model presented in Figure 3.2 (see Chapter 3), and use the criteria listed in Table 8.1 (see Chapter 8) according to the relative importance as discussed in Chapter 4. An effective assessment of GR can be achieved using the method *ScoreGR* defined in Chapter 5, accounting for the insights discussed in Chapter 8.

9.2 Outlook

Future research in this field should investigate the importance of the criteria excluded by the first study, and determine how the criteria not included in the second study interact with each other. Among the 31 criteria of GR now reported in Table 8.1, only fifteen were involved in the first experiment presented in this dissertation, eight of them were involved in the second experiment, and just five were implemented in the GR assessment method *ScoreGR* evaluated in the third experiment. At the same time, future research should aim to unveil further criteria of GR not yet uncovered, as the answers obtained for the third experiment have provided hints for *planning* being a new criterion of GR.

These studies must be accompanied by a more detailed and comprehensive analysis of the conceptual model of GR presented in Chapter 3. As shown in Chapters 5 and 6, only a minor part of the elements described in the conceptual model were involved in the development of the prototype. Future work should focus on deepening on facets that have been only partially implemented, or not yet explored. These include a user's activity (e.g., Couclelis, 2009; Hirtle et al., 2011), planning (e.g., Abdalla and Frank, 2012), preferences (e.g., Rashid et al., 2002; Raubal and Panov, 2009; Schlieder and Kremer, 2011) and a user's social environment (e.g., Mizzaro and Vassena, 2011), which can lead to a finer description of a user's context. Future work should also deepen on the understanding of geographic entities' affordances, which could be uncovered by mean of semantic enrichment (Alves and Pereira, 2012), including its temporal aspect (Alazzawi et al., 2012). This will be a fundamental step in improving the effectiveness of the topicality criterion estimation. Moreover, the definition of context in mobile information services is still a very active field of research (e.g., Emmanouilidis et al., 2012), and new aspects of context will most likely become evident and possible to compute in the future.

Empirical studies should be thereafter conducted to first investigate the importance of all the listed criteria of GR, and then to analyse the interaction between the criteria studied in this dissertation and the other criteria. For instance, it has been shown that the criterion currency is among the primary criteria of GR, but further experiments will be necessary to investigate its relation with spatio-temporal proximity. A geographic entity might be less relevant because the information available referring to it is out-dated, but this also influences its spatio-temporal proximity. In fact, the estimation of its spatio-temporal accessibility could be unreliable, since it is based on out-dated information, and this could lower the relevance of the geographic entity. It is clear that the criteria discussed in Chapter 4 are not fully independent from each another, and this is even more evident in the case of geographic criteria, because of spatial autocorrelation (O'Sullivan and Unwin, 2003). The study of these dependencies could be of central importance in achieving a better understanding of GR.

Future development in the research topics mentioned above would obviously demand further developments in the GR assessment method. *ScoreGR* has been design to effectively assess GR in the scope of the five criteria discussed in Chapter 5. For instance, these criteria accounted for information about the user's activity, location, and information need, but did not involve any information about a user's social context. Once the role of criteria related to that facet of the user's context is investigated, a related numerical score should be established, and integrated into the score combination model of *ScoreGR*. Additional studies will become necessary to evaluate the effectiveness of the updated GR assessment method. This applies to all the criteria discussed in this dissertation, and to all the criteria not yet uncovered. However, including more criteria in an assessment method, and related experimental evaluation, would result in higher complexity. Therefore, careful attention should be paid to both the selection of further criteria to study and implement, and to the selection of procedures to evaluate the effectiveness of developed methods.

Assuming to be able to effectively assess GR for a selected set of criteria, the main objective of this assessment is to achieve an appropriate adaptation of a mobile geographic information service with respect of a user's activity or geographic information need. The representation of GR within the user interface of the same service will then play a fundamental role. An inappropriate representation of the relevance of the entities can mislead the user, and make a service very difficult to use. Crease and Reichenbacher (2011) argued that the map representation in such mobile applications should be carefully designed, targeting the specific task the user is assumed to perform. This should lower the cognitive load, and allow the user to focus on the relevant information, and not to be distracted from the surrounding environment.

Finally, there are interesting lines of research spreading from the concept of GR towards other research fields. The conceptual model of GR presented in Chapter 3 includes the user's preferences and social environment as part of the context in which the information seeking process takes place. The arguments discussed in Chapter 8 call for integrating the user's planning process into the relevance assessment. This suggests that a research field focusing on GR should strongly interact with the fields of social network analysis and personal information management. It is also clear that a GR-based mobile information service implemented in a deployed application cannot avoid to account for the most recent advances in terms of privacy (Hashem and Kulik, 2011, e.g.). In fact, such a service would deal with sensitive data, which have to be handled properly. The perspective for this line of research on GR should be to join the fields mentioned above, together with the fields of MIR (and therefore IR and GIR), LBS, recommender systems, context-aware computing, and mobile cartography, in a common ground of research on the conceptual and computational development of future mobile information services.

Appendix A

Material for Experiment I

The following sections describe the material used in Experiment I[1] (see Section 4.2).

A.1 Questionnaire statements

In the main section of the questionnaire, the participants were asked whether they agree or disagree (on a 7-point Likert scale) with the 15 statements presented in Table A.1. Each statements represents one of the criteria taken into account.

A.2 Questionnaire structure

A total number of 132 participants took part in this experiment. A first group of 53 participants was gathered by sending e-mails to different research mailing-lists, but also to groups of colleagues and friends (including researchers in Computer Science and Geography, but also non-academics). This first group participated in a web survey developed using the online service SurveyMonkey[2], and will be referred to as "SurveyMonkey survey" (SMs) group. A second group of 39 participants, and a third group of 40 participants was gathered through the online service Amazon Mechanical Turk[3], and they will be referred to as "Amazon Mechanical Turk survey 1" (AMTs1) group, and "Amazon Mechanical Turk survey 2" (AMTs2) group, respectively.

The first page of each questionnaire stated the objective of the project and the purposes of the study. Then, participants were asked whether they agree or disagree (on a 7-point Likert scale) with the 15 statements presented in Table A.1. Each statements represents one of the criteria taken into account.

On the second page in the questionnaire presented to the SMs group, the participants were asked about their age and gender, and to state how frequently they use online yellow pages, digital maps, and mobile maps. On the third page, the participants were asked to rank a list of seven general criteria (summarizing the classes of criteria shown in Table

[1]This section is largely based on (De Sabbata and Reichenbacher, 2012).
[2]`http://www.surveymonkey.com`
[3]`https://www.mturk.com/mturk/welcome`

Table A.1: Statements representing the criteria in Experiment I.

Criterion	Questionnaire statement
Appropriateness	A place that offers just the services you need is more relevant than a place that also offers other services.
Coverage	A place that offer all the services you need is more relevant than a place that offers just some of them.
Novelty	A place that was previously unknown to you is more relevant than a place already known.
Availability	The more information available about a place, the higher is the relevance of the place.
Accuracy	The more accurate the information about a place, the higher is the relevance of the place.
Currency	The more current, recent, timely, up-to-date the information about a place, the higher is the relevance of the place.
Dynamism	The more dynamic, active or interactive the presentation of information, the higher is the relevance of the presented place.
Presentation quality	The more the information about a place is presented in a certain format or style, or offers output in a way that is helpful, desirable, or preferable, the higher is its relevance.
Spatio-temporal proximity	It is important to take into account whether the place (or a related event) will be available at the time you will be able to reach it (e.g. whether you can reach the shop before it closes).
Directionality	If you are driving, cycling, or walking, a place on your future path is more relevant than a place already passed.
Visibility	A place that is visible is more relevant than a place that you can not see from your point of view.
Hierarchy	Other things being equal (including distance), a place in the same city or district is more relevant than a place in another one.
Anchor-points proximity	A place that is close to a location you visit frequently (e.g. home or work place) is more relevant than a place in an area you are not used to visit.
Cluster	Other things being equal (including distance), a place close to a group of similar places (e.g. a shop in a shopping center) is more relevant than an isolated place.
Co-location	Other things being equal (including distance), a place that satisfies common co-location rules (e.g. a hotel with a restaurant nearby) is more relevant than a place that does not satisfy the same co-location rules (e.g. a hotel without a restaurant nearby).

4.1) from the most important to the least important. Pages four and five presented a set of 15 statements (see Table A.1) to participants. The first eight criteria on the fourth page and the remaining seven criteria on the fifth page. On both pages, a brief introduction was used to add some context to the questions. On the last page, an open question gave the opportunity to the respondents to specify further criteria, that they would use to judge the relevance of a geographic entity, and to give us any further comment or suggestions.

In the questionnaire presented to the ATMs1 group, the 15 statements were presented to the participant at once on the second page (i.e., all the three classes in Table A.1). The questionnaire presented to the ATMs2 group used a slightly modified structure, that was set up in order to better fit the style commonly used in Amazon Mechanical Turk. The statements were presented to the participant on three different pages (i.e., one for each class in Table A.1). In both cases, on the last page, an open question gave the opportunity to the respondents to specify not mentioned criteria that they would use to judge the relevance of a geographic entity, and to give us any further comments or suggestions.

Appendix B

Material for Experiment II

The following sections describe the material used in Experiment II[1] (see Section 4.3). A total of 110 participants took part in this experiment, gathered by sending e-mails to different mailing-lists, Google Groups[2], and Yahoo Groups[3]. Participant were randomly assigned to one of the four sub-scenarios described below.

B.1 Base map

In both scenarios, the base map (see Figures B.1 and B.2) was derived from the geometries available on OpenStreetMaps[4] for the town of Gorizia (Friuli-Venezia Giulia, Italy), assuming that most of the participants would not be familiar with this town and thus avoid a recognition effect. The base map was flipped vertically, the city center limits have been chosen arbitrarily (i.e., they do not reflect the actual boundaries of the town center of Gorizia), distinguishing buildings have been reshaped, and some park area have been arbitrarily added. None of the entities added to the maps (i.e., hotels, restaurants, museums, and tourist attractions) represent real entities in Gorizia. The three photos used in the first scenario do refer to hotels and bed&breakfasts in Gorizia, but they have been arbitrarily chosen from the images obtained by searching 'hotels Gorizia' via Google Images[5], and arbitrarily assigned to entities on the map that do not represent existing hotels in Gorizia. The reported prices and opening hours have also been arbitrarily chosen, but are based on plausible values.

B.2 Scenario 1

In scenario 1, participants were presented with the following situation:

'*Consider the following scenario. You are visiting a city you have never been before.*

[1]This section is largely based on (De Sabbata and Reichenbacher, 2012).
[2]`http://groups.google.com`
[3]`http://groups.yahoo.com`
[4]`http://www.openstreetmap.org`
[5]`http://images.google.com`

Just after arrival you visit one of the museums in the city center. After the museum visit you feel tired and look for a hotel for the night. Your digital city-guide on your mobile device suggests 6 hotels that fit your needs in terms of costs, availability, and offered services. The suggested hotels are all located at about the same distance from your current location. The map below indicates your current position and the 6 suggested hotels, labeled A to F. Please rate each hotel based on your needs described in the above scenario and the available information on the map'.

Sub-Scenario 1 A

A total of 28 out of 110 participants took part to the first sub-scenario (referred to as S1A). In this sub-scenario (see Figure B.1(a)) the position of hotels, museums and restaurants is shown on the map, together with the position of the participant and her previous route. The city center (i.e., the touristic zone) is highlighted in a brownish color, whereas the residential areas are colored in grey. Three hotels are located in the city center: hotel 'C' is located near restaurants, museums and tourist attraction; hotels 'E' and 'F' are located near restaurants, with 'F' being a bit closer to them than 'E'. Three hotels are located in the residential area: hotels 'A' and 'B' are located close to the city center; hotel 'D' is located far away from the city center.

The hypothesis is that participants would use the available information and judge the relevance of the hotels using the criteria *hierarchy* and *co-location*, i.e.: the participant would take into account the distinction between the city center and the peripheral urban areas, where the first is more relevant than the others; the participant would take into account restaurants, museums and tourist attractions, where the hotels near those POI are more relevant than the others.

Sub-Scenario 1 B

A total of 25 out of 110 participants took part to the second sub-scenario (referred to as S1B). In this sub-scenario (see Figure B.1(b)) the position of museums and restaurants is shown on the map, together with the position of the participant and her previous route. The position of hotels is also displayed, and in some cases it is accompanied by further information on the price of the room or a hotel picture. The hotels are placed in the same position as they were placed in sub-scenario S1A. Detailed price information, and a picture have been attached to hotel 'E' (located in the city center, quite close to some restaurants), and to hotel 'D' (located in the residential area, far away from the city center). Detailed price information has been attached to hotel 'B' and a picture has been attached to hotel 'A', which are close to each other, just outside of the city center. No further information has been attached to the remaining two hotels.

The hypothesis is that participants would use the available information and judge the relevance of the hotels using the two criteria mentioned in the first sub-scenario (S1A), along with the criteria *accuracy*, *availability*, and *presentation quality*, i.e.: the participant would take into account the accuracy of the information about the price, where hotels with detailed price information are more relevant than the others; the participant would take into account the availability of information, where the hotels presenting information about the price are more relevant than the others; the participant would take into account the quality of the presentation, where the hotels presenting an image (that is, presenting information about the hotel in a way that is common to be found in touring guides and websites) are more relevant than the others.

B.3 Scenario 2

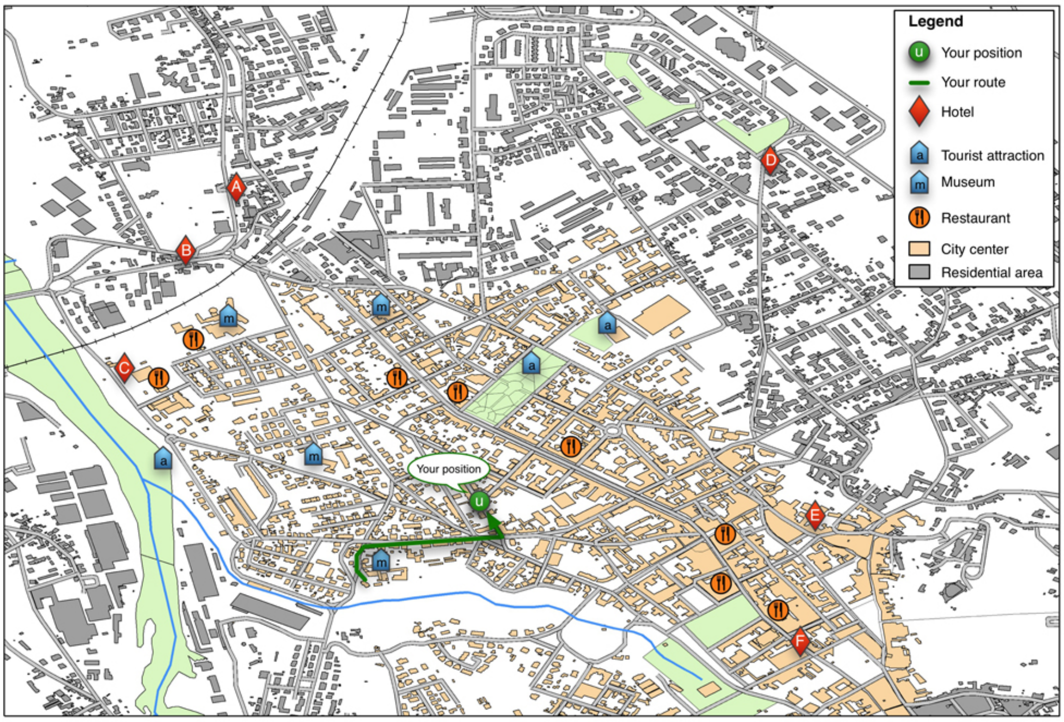
In scenario 2, participants were presented with the following situation:

'*You are visiting a city you have never been before. Just after arrival in the early morning you visit one of the museums in the city center. At 13:45H you are hungry and decide to have a late lunch. Your digital city-guide on your mobile device suggests 9 possible restaurants that fit your needs in terms of cost and offered dishes. The suggested restaurants are all located at about the same distance from your current location: a 10 minute walk. You have not booked a table at any of those restaurants and you do not know anything about table availability either. The map below indicates your current position and the 9 suggested restaurants, labeled A to I, including their opening hours. Please rate each restaurant based on your needs described in the above scenario and the available information on the map*'.

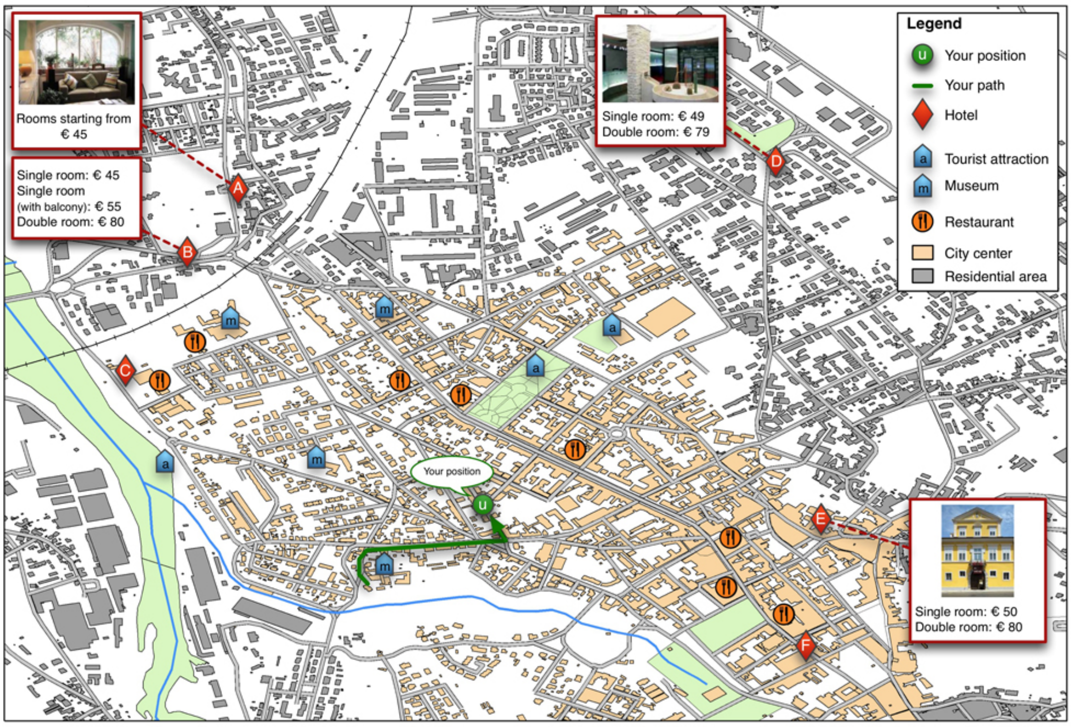
Sub-Scenario 2 A

A total of 28 out of 110 participants took part to the first sub-scenario (referred to as S2A). In this sub-scenario (see Figure B.2(a)) the position and opening hours of the restaurants are shown on the map, together with the position of the participant and the current time. The city center (i.e., the touristic zone) is highlighted in a brownish color, whereas the residential area are colored in grey. Seven restaurants are located in the city center. Three of them have been placed in order to form a cluster (these are the restaurants 'E', 'F', and 'G'), the other four restaurants have been placed in function of their role in the second sub-scenario (as explained in the next paragraph). Two more restaurants are located in the residential area, close to each other.

The hypothesis is that participants would use the available information and judge the relevance of the restaurants using the criteria *spatio-temporal proximity*, *hierarchy*, and *cluster*, i.e.: the participant would take into account the opening hours of the restaurants, that is that the restaurants 'C' (today closed) and 'H' (will close 5 minutes after she could arrive there) would not be relevant; the participant would take into account the visible distinction between the city centre and the peripheral urban areas, where the first is more relevant than the others; the participant would take into account the visible clusters, where the restaurants that are part of a cluster would be more relevant than the others (if she does not find a place in one she can try in the others nearby).

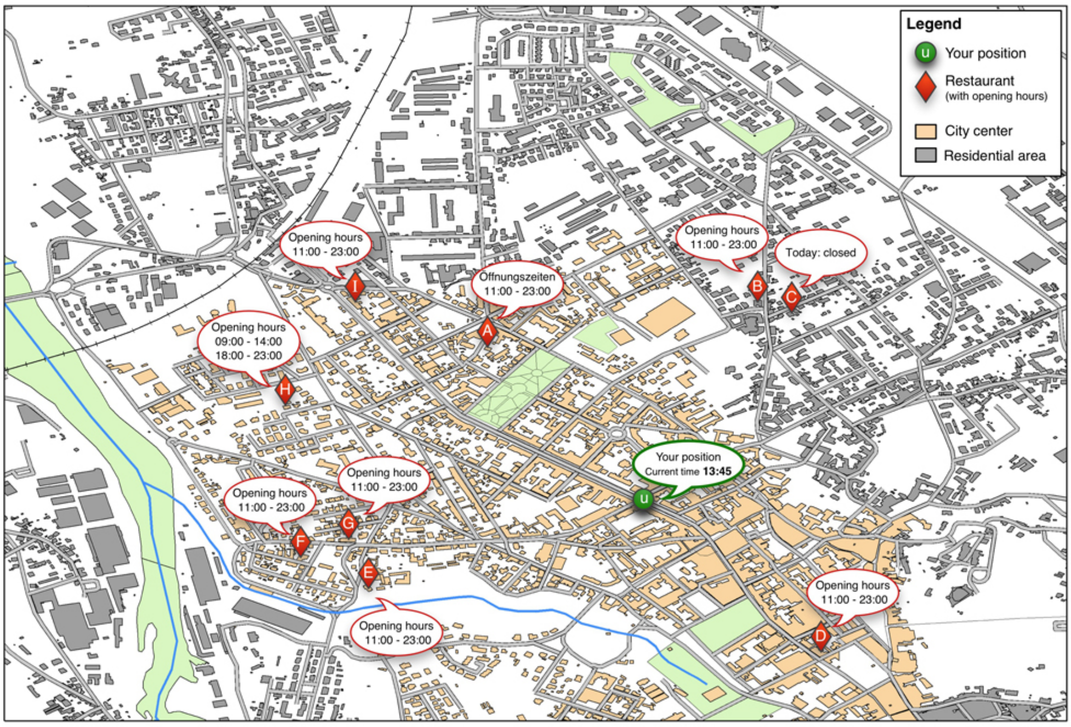


(a)

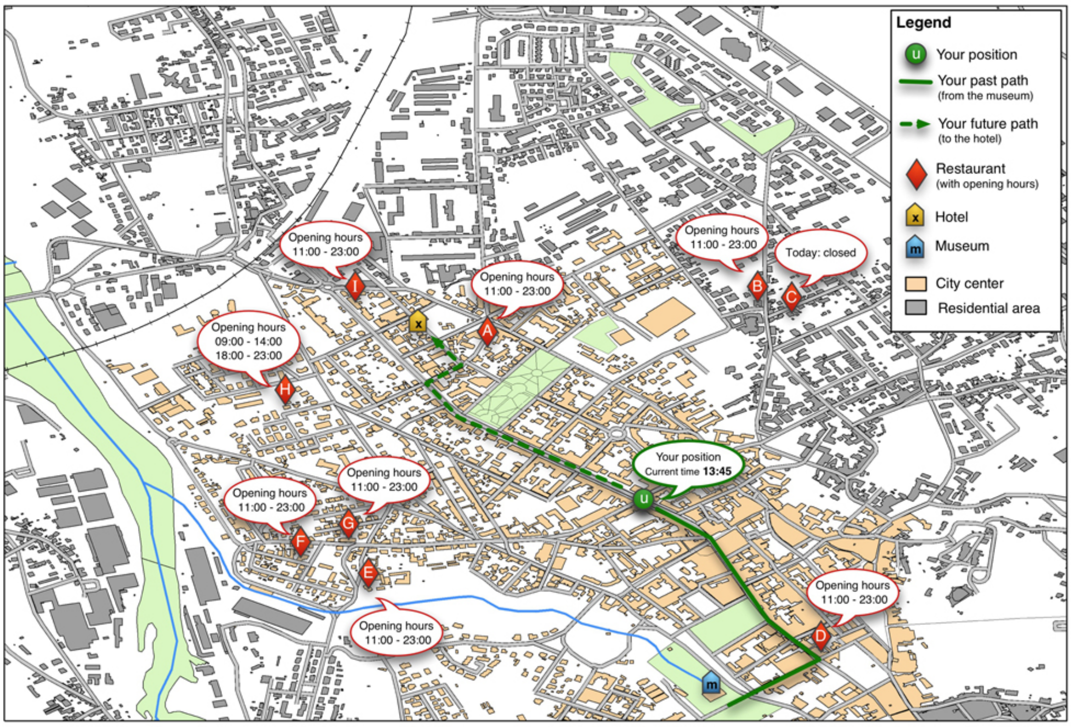


(b)

Figure B.1: Scenario 1: sub-scenario S1A (a) and sub-scenario S1B (b).



(a)



(b)

Figure B.2: Scenario 2: sub-scenario S2A (a) and sub-scenario S2B (b).

Sub-Scenario 2 B

A total of 29 out of 110 participants took part to the second sub-scenario (referred to as S2B). In this sub-scenario (see Figure B.2(b)) the position and opening hours of the restaurants are shown on the map. The participant knows the path she has cone from the museum to her current location (which is displayed on the map), and she knows that, after lunch, she has to go back to the hotel (which is displayed on the map) in order to be able to pack her stuff and leave in the early afternoon. The restaurants are located in the same position as they were placed in sub-scenario S2A. Restaurant ‘A’ is located very close to the hotel and very close the participant’s planned path, whereas restaurant ‘I’ is located at the same distance, but on the other side. Restaurant ‘D’ is located on the participant’s past path, in the opposite direction with respect to the participant’s future path.

The hypothesis is that participants would use the available information and judge the relevance of the hotels using the three criteria mentioned in the first sub-scenario, along with the criterion *directionality*, i.e.: the participant would take into account her direction and destination — the less one has to divert from the shortest path to the hotel, the higher the relevance of the restaurant.

B.4 Questionnaire statements

In the final part of the questionnaire, the participants were asked whether they used the hypothesized criteria. The participant was presented with the sentences reported in Table B.1. An optional comment box was also provided.

Table B.1: Sentences representing the hypothesized criteria.

Criterion	Statement	Sub-scenarios
hierarchy	I have made into account the distinction between the city center and the peripherial urban areas, where the first is more relevant then the others.	S1A, S1B, S2A, S2B
co-location	I have taken into account the restaurants, museums and tourist attractions, where the hotels near those POI are more relevant then the others.	S1A, S1B
availability	I have taken into account the availability of information, where the hotels presenting information about the price are more relevant then the others.	S1B
accuracy	I have taken into account the accuracy of the information about the price, where the hotels with detailed information on the price are more relevant then the others.	S1B
presentation quality	I have taken into account the quality of the presentation, where the hotels presenting an image are more relevant then the others.	S1B
spatio-temporal proximity	I have taken into account the opening hours of the restaurants, that is that the restaurants "c" (Today closed) and "h" (that will close 5 min after she could arrive there) are not relevant.	S2A, S2B
cluster	I have taken into account the groups of restaurants, where the restaurants with other restaurants nearby are more relevant then the others (e.g., if I do not find a place in one I can try in the others nearby).	S2A, S2B
directionality	I have taken into account my direction and future destination, the less I have to divert from the shortest path to the hotel, the higher the relevance of the restaurant.	S2B

Appendix C

Material for Experiment III

The following sections describe the material used in Experiment III (see Chapter 7). The GR assessment methods referred below are described in Chapters 5 and 6.

C.1 Scenario 1

The first scenario is located in the Chamartín district of Madrid, in particular in the neighborhoods Prosperidad and Ciudad Jardín. This area has been selected because it includes a cluster of supermarkets, identified nearby the Calle del Corazón de María. This provides the opportunity to evaluate the role of the criterion cluster in the case of a category of entities which do not commonly cluster. The area also includes many pharmacies, which are used to evaluate the role of the criterion co-location, taking into account the third co-location rule in Table C.2. In order to investigate the role of the criterion spatio-temporal proximity, the opening hours of one of the supermarkets have been set to closing before the user could reach it, and the opening hours of another supermarket have been set so that the user would have less time to spend there than required.

The base maps (see Figures C.1 and C.2) were derived from the geometries available on OpenStreetMaps[1], assuming that most participant would not recognise it when displayed at large scale (i.e., $1 : 8'000$), without street names, having rotated the base data by 105°counterclockwise.The participants were presented with the following situation:

'*It is 19.15, you finish work, and you are outside of your workplace. You are going home, but you need to buy some stuff at the supermarket, including bread, vegetables, and a couple of dental care articles. You have to be home by 20.20, thus you have about one hour and 5 minutes. Shopping at the supermarket will take you at least 20 minutes, which leaves you with a maximum of 45 minutes to walk to the supermarket and then home. You do not want to walk further then needed, and the more time you can spend at the supermarket the better (you do not want to shop in a hurry). Since you are not sure you will find the dental care articles you want at the supermarket, you would need*

[1] `http://www.openstreetmap.org`

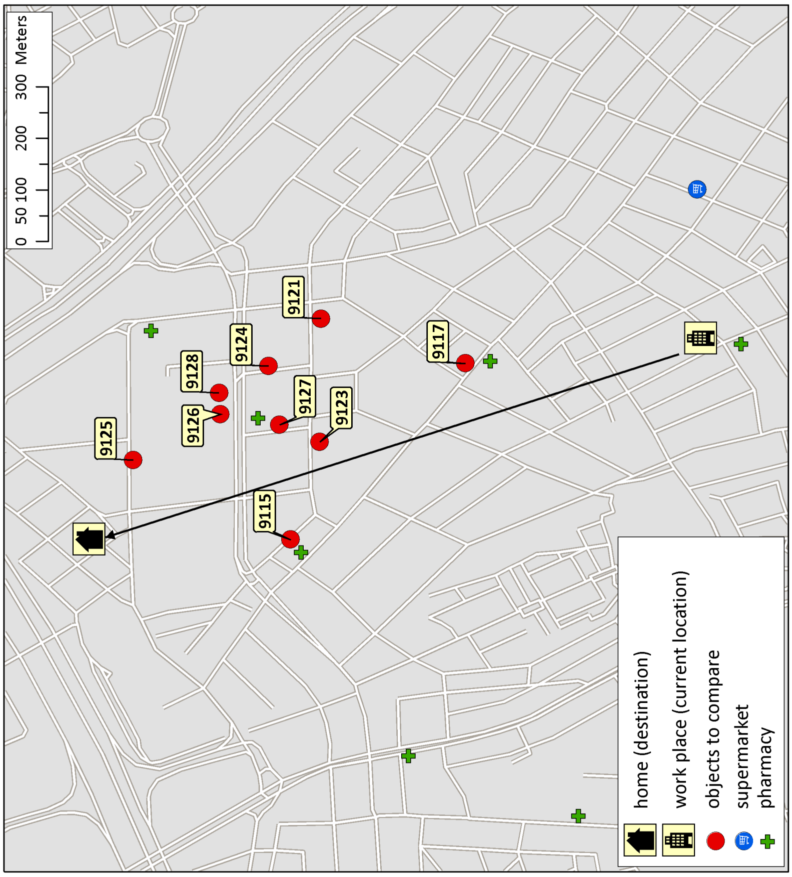


Figure C.1: Overview of Scenario 1.

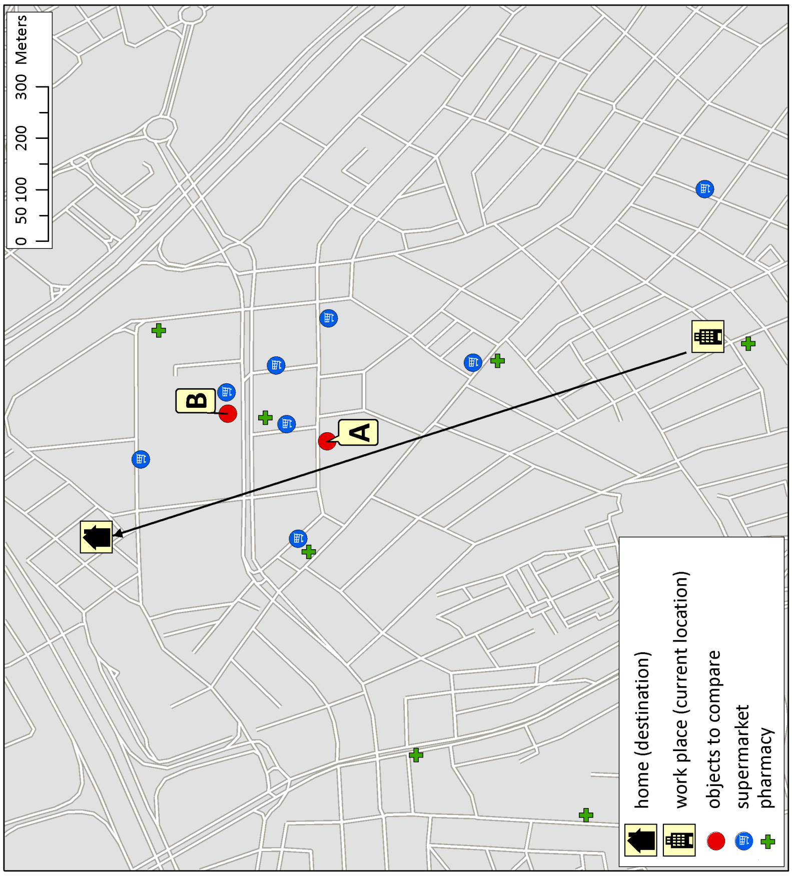


Figure C.2: Example of map presented to the participants of Scenario 1.

Table C.1: Data for entities in Scenario 1 (in the table SM* refers to the cateogry "supermarket").

	9115	9117	9121	9123	9124	9125	9126	9127	9128
Category	SM*	SM*	SM*	SM*	SM*	SM*	SM*	SM*	SM*
Travel distance	993 m	979 m	1104 m	960 m	1059 m	1030 m	1021 m	980 m	1054 m
Closes at	22:00	20:00	20:00	19:00	20:00	20:00	20:00	20:00	22:00
Time left	48′	39′	35′	0′	34′	30′	33′	34′	47′
Angular deviation	8°	13°	21°	4°	15°	6°	10°	7°	12°
Cluster									
cardinality	0	0	0	5	5	0	5	5	5
distance	204 m	292 m	136 m	84 m	109 m	189 m	41 m	84 m	41 m
Co-location									
cardinality	1	1	0	1	2	1	1	1	2
distance	16 m	36 m	201 m	110 m	98 m	194 m	58 m	37 m	80 m
Ranks									
Baseline I	4	2	9	1	8	6	5	3	7
Baseline II	4	2	9	1	8	6	5	3	7
GR	2	3	7	–	5	8	6	4	1
Prob. GR	1	3	5	–	8	6	7	4	2

Scenario 1: '*It is 19.15, you finish work, and you are outside of your workplace. You are going home, but you need to buy some stuff at the supermarket, including bread, vegetables, and a couple of dental care articles. You have to be home by 20.20, thus you have about one hour and 5 minutes. Shopping at the supermarket will take you at least 20 minutes, which leaves you with a maximum of 45 minutes to walk to the supermarket and then home. You do not want to walk further then needed, and the more time you can spend at the supermarket the better (you do not want to shop in a hurry). Since you are not sure you will find the dental care articles you want at the supermarket, you would need to go to a pharmacy. If possible, you would prefer to do your shopping just in one place. Only if you do not find the articles you need in the supermarket will you need to visit the pharmacy. Thus, you are searching for a supermarket with at least one pharmacy nearby. Since you are not sure whether you will still find fresh bread in the first supermarket you will visit, supermarkets with other supermarkets nearby are preferable*'.

Table C.2: Co-location rules (see Section 6.4) taken into account in Experiment III.

Premise	Conclusion	Instances	Probability	Max. Cardinality
hotel	pharmacy	83	36%	5
hotel	restaurant	355	48%	18
supermarket	pharmacy	154	38%	5
restaurant	pub	963	37%	18

to go to a pharmacy. If possible, you would prefer to do your shopping just in one place. Only if you do not find the articles you need in the supermarket will you need to visit the pharmacy. Thus, you are searching for a supermarket with at least one pharmacy nearby. Since you are not sure whether you will still find fresh bread in the first supermarket you will visit, supermarkets with other supermarkets nearby are preferable'.

The prototype described in Chapter 6 has been used to assess GR, and to run the baseline methods for the scenario described above. Aiming to perform the evaluation on about a dozen entities, and given the calculated scores, the top-8 entities from all the four calculated ranks have been taken into account. A total of 9 entities have been selected, whose rank differs considerably from one assessment method to the other. Information about the selected entities and their obtained ranks are presented in Table C.1.

C.2 Scenario 2

The second scenario is located in Centro district of Madrid, in particular in the neighborhood Sol. This area has been selected because of the large number of hotels, along with the numerous restaurants and pharmacies. The presence of entities belonging to those categories can be used to studied the role of the criterion co-location, taking into account the first and the second co-location rules in Table C.2. The area also include two clusters of hotels, which are exploited to evaluate the role of the criterion cluster in the case of this category of entities.

The base maps (see Figures C.3 and C.4) were derived from the geometries available on OpenStreetMaps[2], assuming that most participant would not recognise it when displayed at large scale (i.e., $1:5'000$), without street names, having rotated the base data by 105°counterclockwise, and excluded the area of the Royal Palace from the map. The participants were presented with the following situation:

'*You have arrived for a conference in a city you have never been to before. You did not manage to book a hotel. The first morning session finishes and you have one hour before the next session starts. During this hour you want to find a hotel to spend a couple of nights. The closer the hotel is to the conference place, the better. Moreover you would prefer to find a hotel with some restaurants nearby so that you can have dinner in the*

[2] `http://www.openstreetmap.org`

evening. Moreover you would prefer to have a pharmacy nearby your hotel, since it is allergy season and you may need medications for your allergy. Your map displays the location of the hotels, along with the position of the restaurants, pharmacies, and the conference location (your current location), but you do not know whether those hotels have available places. Since you do not have much time, you prefer to go to hotels with other hotels nearby, so that you have other opportunities in case the first hotel you visit is fully booked. You do not have specific preferences about the type of hotel or services for each of the presented hotels'.

The prototype described in Chapter 6 has been used to assess GR and to run the baseline methods for the scenario described above. Aiming to perform the evaluation on about a dozen entities, and given the calculated scores, the top-4 entities from all the four calculated ranks have been taken into account. A total of 10 entities have been selected, whose rank differs considerably from one assessment method to the other. Information about the selected entities and their obtained ranks are presented in Table C.3.

C.3 Scenario 3

The second scenario is located in what actually is the Centro district of Madrid, in particular in the neighborhood of Sol (the same area used in the second scenario, but a different centre and orientation). This area has been selected because of the large number of restaurants, along with the numerous pubs. The presence of entities belonging to those categories can be used to studied the role of the criterion co-location, taking into account the fourth co-location rule in Table C.2. The area also include many clusters of restaurants, which are exploited to evaluate the role of the criterion cluster in the case of this category of entities.

The base maps (see Figures C.5 and C.6) were derived from the geometries available on OpenStreetMaps[3], assuming that most participant would not recognise it when displayed at large scale (i.e., $1 : 5'000$), without street names, having rotated the base data by 60°clockwise, and excluded the area of the Royal Palace from the map. The participants were presented with the following situation:

'*You and your friends are planning to go and eat something together after work has finished, and go for a drink later to celebrate your birthday. Within three hours, all of you have to reach the bus station to take the bus home. You are searching for a suitable restaurant to go to. The shorter the distance from your current position and then to the bus station, the better – since you would then have more time to spend at the restaurant. Moreover, since you did not book a table and you do not know whether you will find a place in the first restaurant you will visit, you prefer restaurants with other restaurants*

[3] `http://www.openstreetmap.org`

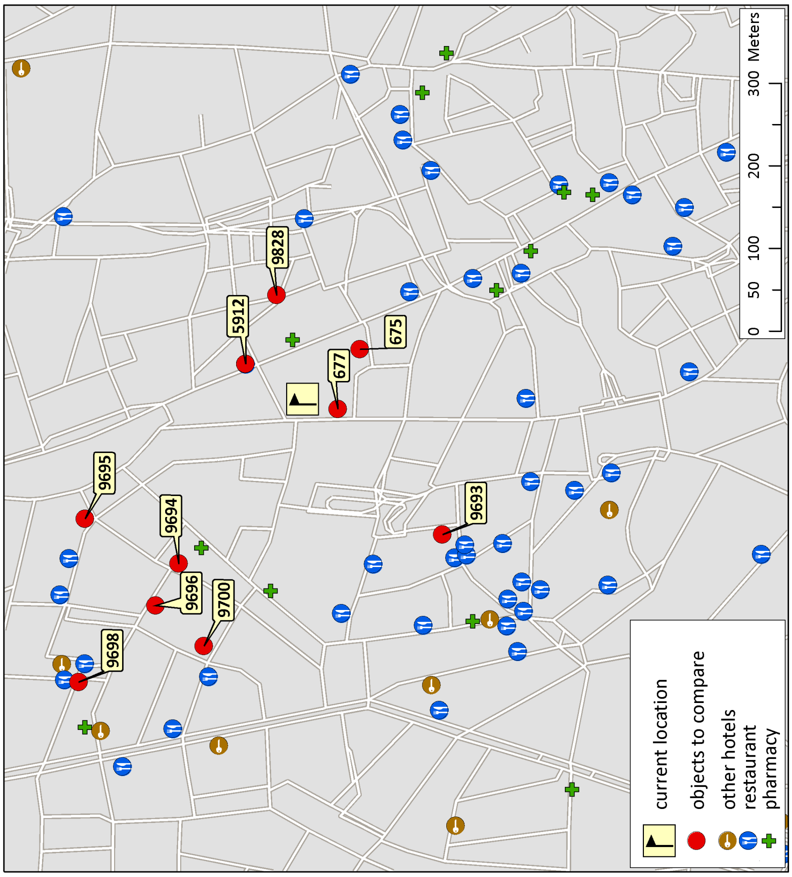


Figure C.3: Overview of Scenario 2.

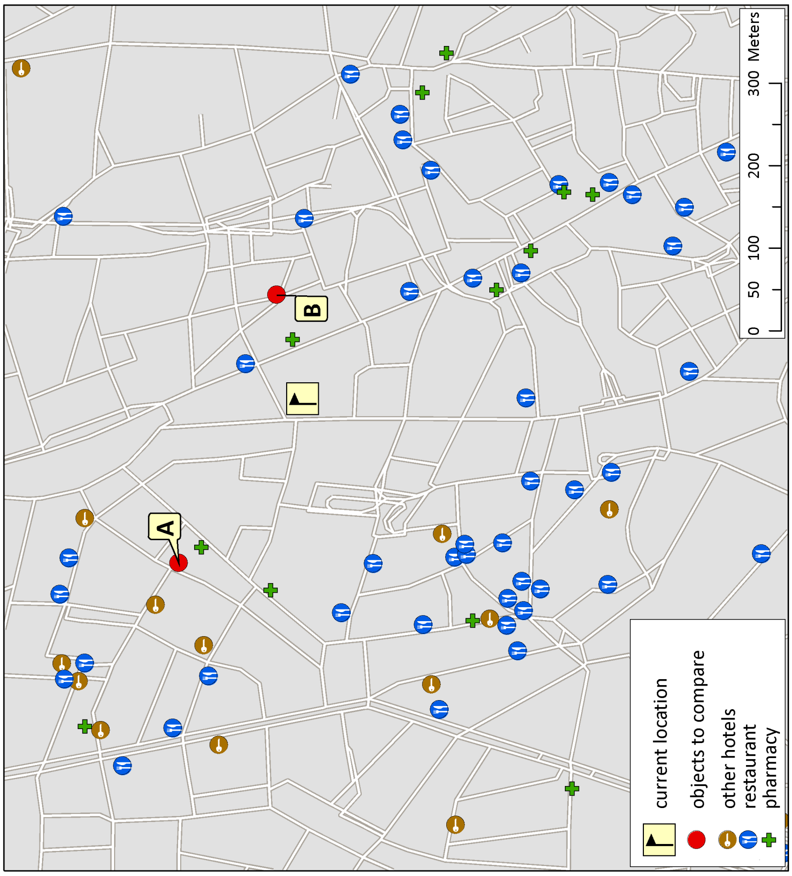


Figure C.4: Example of map presented to the participants of Scenario 2.

Table C.3: Data for entities in Scenario 2.

	675	677	5912	9693	9694	9695	9696	9698	9700	9828
Category	pub	nightclub	restaurant	hotel	hotel	hotel	hotel	hotel	hotel	hotel
Travel distance	149 m	78 m	140 m	449 m	431 m	468 m	533 m	749 m	596 m	255 m
Cluster										
cardinality	–	–	–	0	3	0	3	3	3	0
distance	–	–	–	118 m	57 m	125 m	57 m	29 m	7 m	345 m
Co-location (restaurants)										
cardinality	–	–	–	18	8	4	8	7	9	5
distance	–	–	–	26 m	105 m	47 m	90 m	14 m	37 m	83 m
Co-location (pharmacies)										
cardinality	–	–	–	2	2	1	3	4	3	1
distance	–	–	–	111 m	27 m	120 m	76 m	55 m	85 m	57 m
Ranks										
Baseline I	–	–	–	3	2	4	5	10	6	1
Baseline II	4	1	3	7	6	10	14	21	16	2
GR	206	193	77	6	1	8	4	2	3	7
Prob. GR	51	41	40	3	2	4	5	10	6	1

Scenario 2: '*You have arrived for a conference in a city you have never been to before. You did not manage to book a hotel. The first morning session finishes and you have one hour before the next session starts. During this hour you want to find a hotel to spend a couple of nights. The closer the hotel is to the conference place, the better. Moreover you would prefer to find a hotel with some restaurants nearby so that you can have dinner in the evening. Moreover you would prefer to have a pharmacy nearby your hotel, since it is allergy season and you may need medications for your allergy. Your map displays the location of the hotels, along with the position of the restaurants, pharmacies, and the conference location (your current location), but you do not know whether those hotels have available places. Since you do not have much time, you prefer to go to hotels with other hotels nearby, so that you have other opportunities in case the first hotel you visit is fully booked. You do not have specific preferences about the type of hotel or services for each of the presented hotels*'.

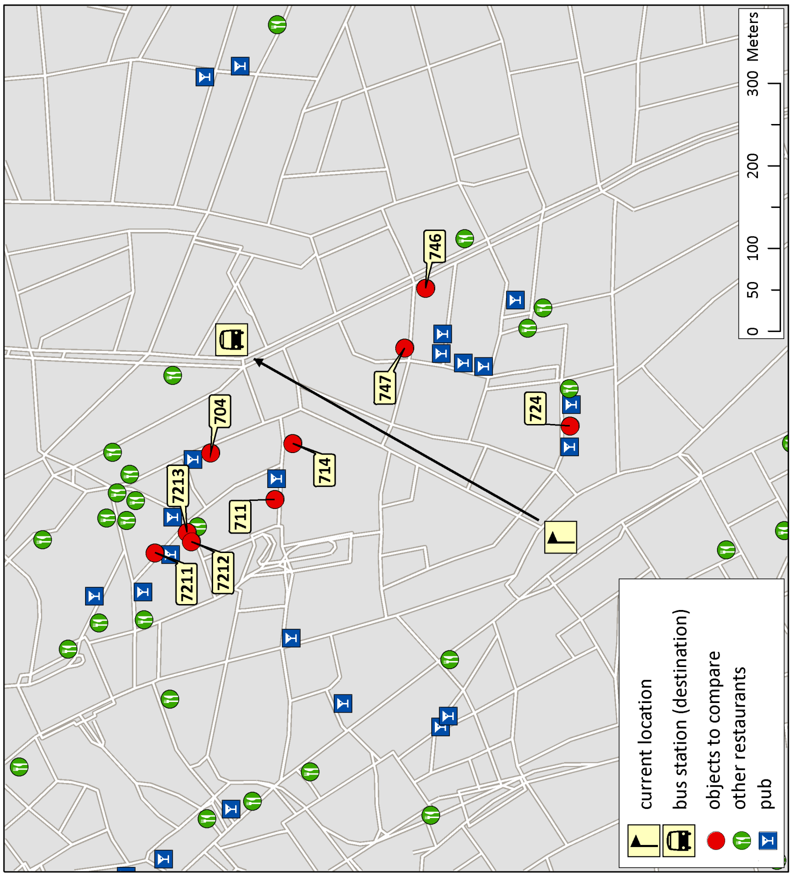


Figure C.5: Overview of Scenario 3.

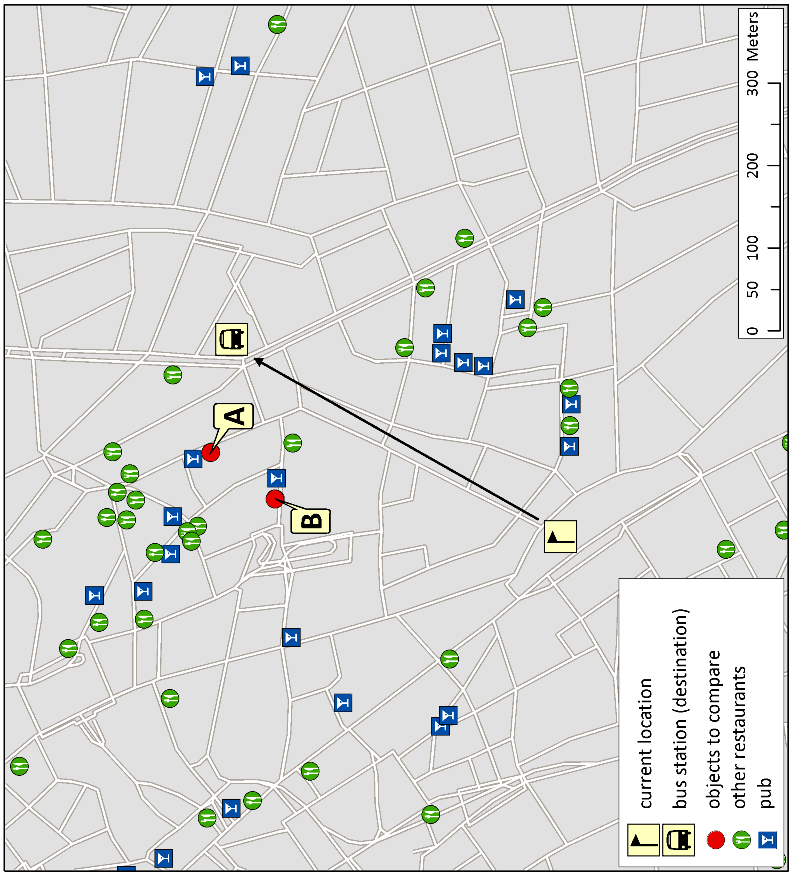


Figure C.6: Example of map presented to the participants of Scenario 3.

Table C.4: Data for entities in Scenario 3 (in the table RT* refers to the cateogry "restaurant").

	704	711	714	724	746	747	7211	7212	7213
Category	RT*	RT*	RT*	RT*	RT*	RT*	RT*	RT*	RT*
Travel distance	508 m	493 m	453 m	493 m	530 m	467 m	640 m	592 m	588 m
Open or closed	open	closed	open	open	open	open	open	open	open
Angular deviation	17°	23°	11°	63°	31°	20°	33°	31°	30°
Cluster									
cardinality	19	19	19	4	3	3	19	19	19
distance	90 m	71 m	71 m	46 m	76 m	77 m	46 m	13 m	13 m
Co-location									
cardinality	6	7	9	7	5	7	7	7	7
distance	17 m	24 m	39 m	23 m	58 m	38 m	16 m	22 m	25 m
Ranks									
Baseline I	5	4	1	3	7	2	19	13	12
Baseline II	5	4	1	3	7	2	19	13	12
GR	1	–	2	38	17	15	3	5	4
Prob. GR	3	–	1	4	5	2	20	13	9

Scenario 3: '*You and your friends are planning to go and eat something together after work has finished, and go for a drink later to celebrate your birthday. Within three hours, all of you have to reach the bus station to take the bus home. You are searching for a suitable restaurant to go to. The shorter the distance from your current position and then to the bus station, the better – since you would then have more time to spend at the restaurant. Moreover, since you did not book a table and you do not know whether you will find a place in the first restaurant you will visit, you prefer restaurants with other restaurants nearby. In choosing the restaurant, you prefer restaurants with pubs nearby, so that you do not need to walk for long to reach the pub. The closer the pubs, the better. The more pubs nearby the restaurant, the better. You do not have specific preferences about the type of restaurant or cuisine*'.

nearby. In choosing the restaurant, you prefer restaurants with pubs nearby, so that you do not need to walk for long to reach the pub. The closer the pubs, the better. The more pubs nearby the restaurant, the better. You do not have specific preferences about the type of restaurant or cuisine'.

The prototype described in Chapter 6 has been used to assess GR and to run the baseline methods for the scenario described above. Aiming to perform the evaluation on about a dozen entities, and given the calculated scores, the top-5 entities from all the four calculated ranks have been taken into account. A total of 9 entities have been selected, whose rank differs considerably from one assessment method to the other. Information about the selected entities and their obtained ranks are presented in Table C.4.

C.4 Crowdsourced judgements

The following pages present a summary of the ordering process performed through the crowdsourcing service CrowdFlower[4]. Each iteration is presented to the participant as a task, introduced with a description of the overall aim of the research, and the specific scenario. This is followed by a list of comparisons to be performed. Each comparison is presented to the participant as shown in Figure C.7. The participants were asked to take into account all available information in order to judged which of the two geographic entities would better fit the user's needs described in the scenario. On the right most side in Figure C.7, a map illustrates the geographic entities that need to be judge, including the location and destination of the user, and the entities of the types involved in the criteria clusters and co-location rules. On the left most side in Figure C.7, the two entities are described. The same information is reported in a tabular form below.

Below the table the answer block is presented, which is divided in three parts. First, participants are asked to choose between the entities. When calculating the overall outcome of a comparison, one point is given to each entity of the pair for each participant who voted for that entity. Participants were also able to classify the two entities as equally relevant, or as both non-relevant. No point is assigned if the participant specified one of the last two options. The difference of the points obtained by the two entities defines which one is more relevant (i.e., the one with more points). The entities are considered equally relevant, if the difference is equal or lower than the 5% of the scores, or if over 80% of the participants specified that the entities are equally relevant. Second, participants are then able to specify whether one or both entities are irrelevant. An entity is considered irrelevant if the related option has been selected by at least half of the participants. Finally, participants are asked to provide a textual motivation for their judgement. All the fields were set as mandatory.

For each iteration, the results have been always calculated twice. In the first calculation, the results have been calculated excluding all the answers given by workers who did not provide an explanation of their judgments (these have been reported as "ma-

[4] `https://crowdflower.com`, last accessed September 2012

Question 5

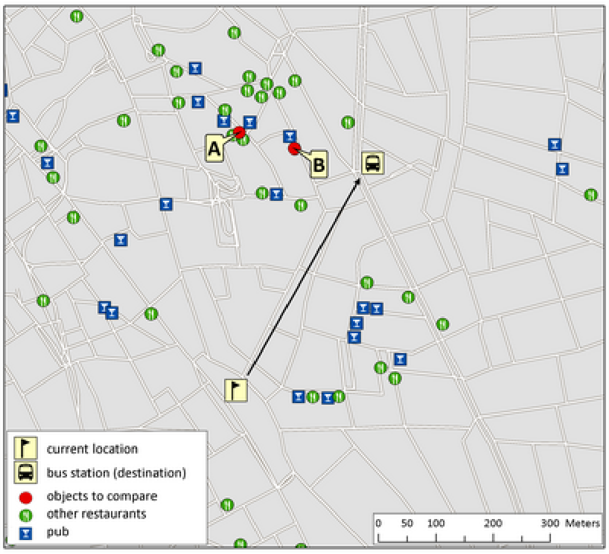
For the scenario described above, compare the two objects described below and shown on the map.

B is a **restaurant**, which is currently **open**. It would take you **8 minutes** to walk from your location to B and then to the bus station. There are other **18 restaurants** nearby and the distance to the closest one is at **2 minutes' walk**. There are **6 pubs** nearby and the distance to the closest pub is **1 minutes' walk**.

A is a **restaurant**, which is currently **open**. It would take you **10 minutes** to walk from your location to A and then to the bus station. There are other **18 restaurants** nearby and the distance to the closest one is at **1 minutes' walk**. There are **7 pubs** nearby and the distance to the closest pub is **1 minutes' walk**.

The information about the two objects is summarised in the table below.

Given this information, which one of the two objects described below and shown on the map better fits your needs?



	Total walking distance	Open or closed	Other restaurants nearby	Closest other restaurant	Pubs nearby	Closest pub
B	8 minutes' walk	open	18	2 minutes' walk	6	1 minutes' walk
A	10 minutes' walk	open	18	1 minutes' walk	7	1 minutes' walk

Answer

- B better fits my needs
- A better fits my needs
- they equally fit my needs
- none of the three options above

Moreover: (if applicable)

- the option B does not fit at all
- the option A does not fit at all

An option "does not fit at all" (i.e., is irrelevant) if it does not satisfy the criteria described in the given scenario.

Please carefully explain why you chose the selected answer, this is very important for our research. Why do you think the object you selected better fits your need?

Figure C.7: Screenshot from the crowdsourcing service CrowdFlower: one of the questions in iteration 2, scenario 3, comparing entities 7123 (A) and 704 (B).

licious" workers). In fact, although the text field requesting for an explanation of the judgements had been set as mandatory, some workers inserted just random text (most of these judgements were also incoherent). Typically, each iteration was affected by 2 up to 4 of such cases (i.e., 5% up to 10% of the workers). In the second calculation, the results have been calculated excluding also all the answers given by workers who provided at least one incoherent answer in the iteration (these have been reported as "inattentive" workers). These are workers who gave to different answers to the duplicated questions, or flagged as irrelevant the same entity preferred in the comparison. In ten over the 74 identified cases (and the 2685 collected judgements), it has been possible to manually correct such disattention starting from the unequivocal explanation given by the worker, as reported in Table C.16. In most cases there is no difference between the outcome obtained taking into account all the non-malicious workers and the just the trusted ones (i.e., excluding the inattentive). Further details are reported in Section 7.1.2.

Once the results for an iteration are obtained, the same procedure is applied to the remaining unordered elements of the list. This ordering process is illustrated for the three scenarios in Figures C.8, C.9, and C.10, from the unordered set in the first line to the ranked list on the bottom of each figure. A detailed report of the collected answers for the three scenarios is proposed in Tables C.10, C.12, and C.14. The black numbers enclosed in rounded rectangles represent unordered sets of entities, each one illustrated with its identification number. In each line, red circles highlight the entities selected as pivots for an iteration. The small red numbers refer to related iteration number. Figures C.8, C.9, and C.10 show the entities aligned according to the final order in all lines, in order to facilitate the visual illustration of the process. The red arrows represent the splitting process triggered by the comparisons with the pivot of the line above. The final ordering is reported in the bottom line of each figure. The entities classified as irrelevant are coloured in grey. Tables C.5 to C.9 summarise the subdivision of the participants among the different scenarios and iterations.

Table C.5: Iterations summary of Scenario 1

				Unique workers			
Iteration	Pairs	(Checks)	Comparisons	Total	Malicious	Inattentive	Trusted
1	8	(+3)	484	44	2	15	27
2	3	(+1)	160	40	4	5	31
3	2		88	44	4	3	37
4	2		80	40	3	2	35
5	1		40	40	1	0	39
Total	16	(+4)	832	189	11	22	156

Table C.6: Iterations summary of Scenario 2

				Unique workers			
Iteration	Pairs	(Checks)	Comparisons	Total	Malicious	Inattentive	Trusted
1	9	(+3)	576	48	1	11	36
2	2		84	42	2	1	39
3	2		80	40	2	1	37
4	1		40	40	1	0	39
Total	14	(+3)	780	159	6	13	140

Table C.7: Iterations summary of Scenario 3

				Unique workers			
Iteration	Pairs	(Checks)	Comparisons	Total	Malicious	Inattentive	Trusted
1	8	(+3)	484	44	4	23	17
2	6	(+2)	320	40	3	12	25
3	2		80	40	4	0	36
4	3		426	42	3	1	38
X	1		43	43	0	0	43
Total	20	(+5)	1053	186	12	34	140

Table C.8: Overall iterations summary of Experiment III

				Unique workers			
	Pairs	(Checks)	Comparisons	Total	Malicious	Inattentive	Trusted
Overall	50	(+12)	2685	416	18	69	329

Table C.9: Number of attended iterations and scenarios per worker

	1 iter.	2 iter.s	3 iter.s	4 iter.s	5 iter.s	6 iter.s	7 iter.s	8 iter.s
Scenario 1	175	11	2	0	1	–	–	–
Scenario 2	151	6	1	1	–	–	–	–
Scenario 3	172	10	3	1	0	–	–	–
Overall	320	69	25	5	0	2	0	3

	1 scenario	2 scenarios	3 scenarios
Overall	317	78	21

Table C.10: Scenario 1

Iteration	Non-malicious						Trusted					
1	Pivot		Other	Bal.	Eq.	None	Pivot		Other	Bal.	Eq.	None
	9115	>	9126				9115	>	9126			
	12		26	-14	4	0	7		19	-12	1	0
	9115	>	9127				9115	>	9127			
	10		29	-19	3	0	8		19	-11	0	0
	9115	<	9117				9115	<	9117			
	32		2	30	6	2	22		2	20	2	1
(check)	31		3	28	6	2	22		2	20	2	1
	9115	<	9123	[22 (22) IRR 9123]			9115	<	9123	[16 (15) IRR 9123]		
	38		2	36	1	1	26		0	26	0	1
(check)	37		2	35	1	2	26		0	26	0	1
	9115	>	9128				9115	>	9128			
	0		35	-35	7	0	0		24	-24	3	0
(check)	1		34	-33	7	0	0		24	-24	3	0
	9115	>	9124				9115	>	9124			
	12		25	-13	5	0	9		16	-7	2	0
	9115	<	9121				9115	<	9121			
	31		5	26	4	2	22		3	19	1	1
	9115	<	9125				9115	<	9125			
	33		3	30	3	3	23		2	21	1	1
2	Pivot		Other	Bal.	Eq.	None	Pivot		Other	Bal.	Eq.	None
	9128	<	9127				9128	<	9127			
	29		4	25	1	0	27		4	23	0	0
(check)	27		7	20	0	0	27		4	23	0	0
	9128	<	9124				9128	<	9124			
	30		4	26	0	0	28		3	25	0	0
	9128	<	9126				9128	<	9126			
	27		6	21	1	0	27		4	23	0	0
3	Pivot		Other	Bal.	Eq.	None	Pivot		Other	Bal.	Eq.	None
	9121	>	9117				9121	>	9117			
	4		35	-31	1	0	2		34	-32	1	0
	9121	>	9125				9121	>	9125			
	14		22	-8	4	0	13		20	-7	4	0
4	Pivot		Other	Bal.	Eq.	None	Pivot		Other	Bal.	Eq.	None
	9126	<	9124				9126	<	9124			
	21		10	11	5	1	20		10	10	5	0
	9126	>	9127				9126	>	9127			
	4		27	-23	6	0	3		26	-23	6	0
5	Pivot		Other	Bal.	Eq.	None	Pivot		Other	Bal.	Eq.	None
	9117	<	9125				9117	<	9125			
	26		8	18	5	0	26		8	18	5	0

Note: The notation A < B indicates that A is ranked higher than B. The notation A > B indicates that A is ranked lower than B. The notation A = B indicates that A and B are tied. The notation A ? B indicates controversial comparisons, which need further investigation.

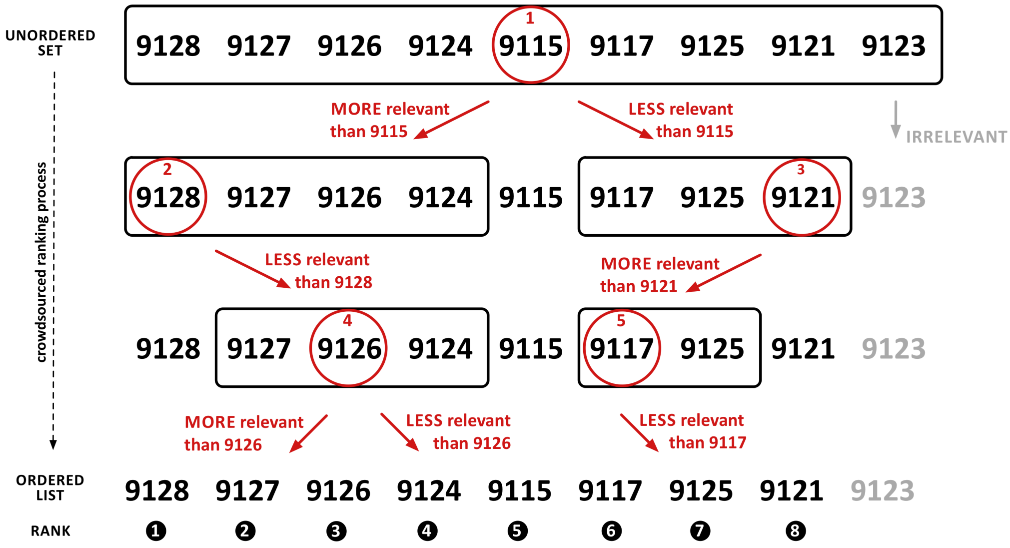


Figure C.8: Illustration of the crowdsourced ranking process for Scenario 1.

Table C.11: Comparison between crowdsourced and computed ranks for Scenario 1.

Entity	*Crowdsourced*	*Baseline1*	*Baseline2*	*ScoreGR*	*GRBM25*
9128	1	7	7	1	2
9127	2	3	3	4	4
9126	3	5	5	6	7
9124	4	8	8	5	8
9115	5	4	4	2	1
9117	6	2	2	3	3
9125	7	6	6	8	6
9121	8	9	9	7	5
9123	IRR	1	1	IRR	IRR
Correlation	Kendall's τ	-.111	-.111	.556*	.333
	p	>.05	>.05	<.05	>.05

* Correlation is significant at the 0.05 level (2-tailed).
** Correlation is significant at the 0.01 level (2-tailed).

Note: the label "IRR" refers to entities identified as irrelevant.

Table C.12: Scenario 2

Iteration	Non-malicious						Trusted					
1	Pivot		Other	Bal.	Eq.	None	Pivot		Other	Bal.	Eq.	None
	9698	<	9693				9698	<	9693			
	28		16	12	4	0	24		8	16	4	0
	9698	>	9700				9698	>	9700			
	1		36	-35	11	0	0		29	-29	7	0
(check)	4		36	-32	8	0	0		29	-29	7	0
	9698	>	9694				9698	>	9694			
	2		43	-41	3	0	0		34	-34	2	0
	9698	<	9695				9698	<	9695			
	38		8	30	2	0	30		5	25	1	0
(check)	37		9	28	2	0	30		5	25	1	0
	9698	>	9696				9698	>	9696			
	4		34	-30	10	0	0		28	-28	8	0
	9698	<	675	[40 IRR 675]			9698	<	675	[29 IRR 675]		
	47		1	46	0	0	36		0	36	0	0
	9698	<	5912	[39 (38) IRR 5912]			9698	<	5912	[28 (27) IRR 5912]		
	46		1	45	0	1	36		0	36	0	0
(check)	45		2	43	0	1	36		0	36	0	0
	9698	<	9828				9698	<	9828			
	31		16	15	1	0	26		10	16	0	0
	9698	<	677	[40 IRR 677]			9698	<	677	[29 IRR 677]		
	47		0	47	0	1	36		0	36	0	0
2	Pivot		Other	Bal.	Eq.	None	Pivot		Other	Bal.	Eq.	None
	9700	=	9696				9700	=	9696			
	10		8	2	22	0	10		8	2	21	0
	9700	>	9694				9700	>	9694			
	12		20	-8	8	0	12		19	-7	8	0
3	Pivot		Other	Bal.	Eq.	None	Pivot		Other	Bal.	Eq.	None
	9695	>	9828				9695	>	9828			
	7		25	-18	4	2	7		24	-17	4	2
	9695	>	9693				9695	>	9693			
	8		26	-18	3	1	7		26	-19	3	1
4	Pivot		Other	Bal.	Eq.	None	Pivot		Other	Bal.	Eq.	None
	9828	>	9693				9828	>	9693			
	10		26	-16	3	0	7		26	-16	3	0

Note: The notation A $<$ B indicates that A is ranked higher than B. The notation A $>$ B indicates that A is ranked lower than B. The notation A $=$ B indicates that A and B are tied. The notation A ? B indicates controversial comparisons, which need further investigation.

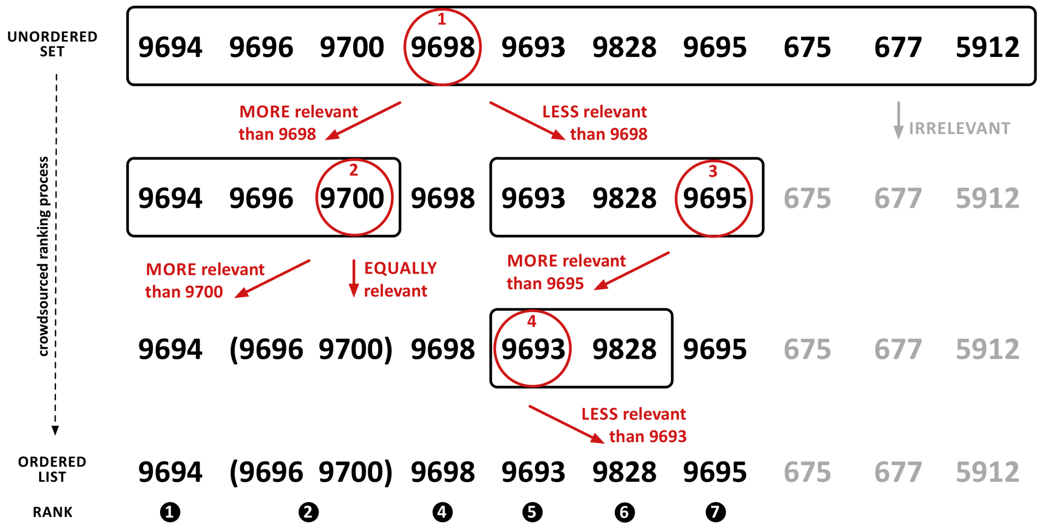


Figure C.9: Illustration of the crowdsourced ranking process for Scenario 2.

Table C.13: Comparison between crowdsourced and computed ranks for Scenario 2.

Entity	*Crowdsourced*	*Baseline1*	*Baseline2*	*ScoreGR*	*GRBM25*
9694	1	2	6	1	2
9696	2	5	14	4	5
9700	2	6	16	3	6
9698	4	10	21	2	10
9693	5	3	7	6	3
9828	6	1	2	7	1
9695	7	4	10	8	4
675	IRR	IRR	4	206	51
677	IRR	IRR	1	193	41
5912	IRR	IRR	3	77	40
Correlation	Kendall's τ	.458	-.442	.861**	.442
	p	>.05	>.05	<.01	>.05

* Correlation is significant at the 0.05 level (2-tailed).
** Correlation is significant at the 0.01 level (2-tailed).

Note: the label “IRR” refers to entities identified as irrelevant.

Table C.14: Scenario 3

Iteration	Non-malicious						Trusted					
1	Pivot		Other	Bal.	Eq.	None	Pivot		Other	Bal.	Eq.	None
(see X)	746	?	711	[9 (12) IRR 711]			746	?	711	[7 (7) IRR 711]		
	16		22	-6	2	0	10		8	2	0	0
(check)	17		21	-4	1	1	10		8	2	0	0
	746	>	747				746	>	747			
	1		23	-22	15	1	1		12	-11	4	1
	746	>	724				746	>	724			
	2		28	-26	9	1	0		14	-14	3	1
(check)	5		25	-20	9	1	0		14	-14	3	1
	746	>	7212				746	>	7212			
	4		33	-29	3	0	3		13	-10	2	0
	746	>	7211				746	>	7211			
	8		30	-22	2	0	5		12	-7	1	0
(check)	12		26	-14	2	0	5		12	-7	1	0
	746	>	7213				746	>	7213			
	2		36	-34	2	0	2		15	-13	1	0
	746	>	704				746	>	704			
	1		38	-37	1	0	1		17	-16	0	0
	746	>	714				746	>	714			
	1		37	-36	2	0	0		17	-17	1	0
2	Pivot		Other	Bal.	Eq.	None	Pivot		Other	Bal.	Eq.	None
	7213	=	7212				7213	=	7212			
	0		1	-1	35	1	0		1	-1	23	0
	7213	<	724				7213	<	724			
	29		8	21	0	0	21		3	18	0	0
(check)	31		5	26	1	0	21		3	18	0	0
	7213	<	7211				7213	<	7211			
	18		0	18	19	0	13		0	13	11	0
(check)	21		1	20	15	0	13		0	13	11	0
	7213	>	714				7213	>	714			
	8		25	-17	4	0	5		17	-12	2	0
(see 4)	7213	?	704				7213	?	704			
	12		18	-6	7	0	9		8	1	7	0
	7213	<	747				7213	<	747			
	32		4	28	1	0	22		1	21	1	0
3	Pivot		Other	Bal.	Eq.	None	Pivot		Other	Bal.	Eq.	None
	724	<	747				724	<	747			
	21		5	16	9	1	21		5	16	9	1
	724	=	7211				724	=	7211			
	17		17	0	2	0	17		17	0	2	0
4	Pivot		Other	Bal.	Eq.	None	Pivot		Other	Bal.	Eq.	None
	704	<	7213				704	<	7213			
	25		8	17	5	0	25		8	17	5	0
	704	>	714				704	>	714			
	5		25	-20	8	0	5		25	-20	8	0
	704	<	7212				704	<	7212			
	18		9	9	11	0	18		9	9	11	0
X	Pivot		Other	Bal.	Eq.	None	Pivot		Other	Bal.	Eq.	None
	746	<	711	[30 IRR 711]			746	<	711	[30 IRR 711]		
	35		5	30	1	2	35		5	30	1	2

Note: The notation A < B indicates that A is ranked higher than B. The notation A > B indicates that A is ranked lower than B. The notation A = B indicates that A and B are tied. The notation A ? B indicates controversial comparisons, which need further investigation.

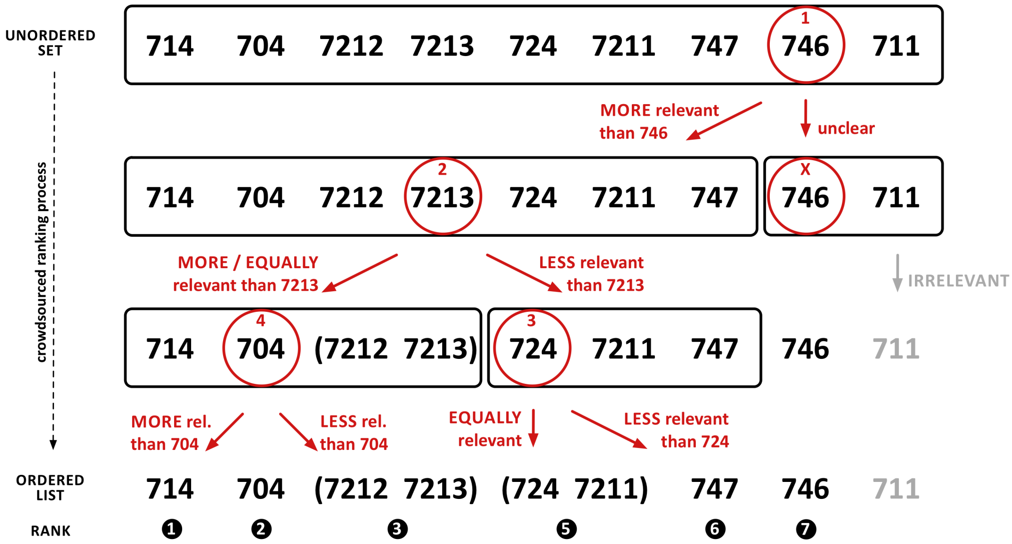


Figure C.10: Illustration of the crowdsourced ranking process for Scenario 3.

Table C.15: Comparison between crowdsourced and computed ranks for Scenario 3.

Entity	*Crowdsourced*	*Baseline1*	*Baseline2*	*ScoreGR*	*GRBM25*
714	1	1	1	2	1
704	2	5	5	1	3
7212	3	13	13	5	13
7213	3	12	12	4	9
724	5	3	3	38	4
7211	5	19	19	3	20
747	7	2	2	15	2
746	8	7	7	17	5
711	IRR	4	4	IRR	IRR
Correlation	Kendall's τ	.057	.057	.686*	.400
	p	>.05	>.05	<.05	>.05

* Correlation is significant at the 0.05 level (2-tailed).
** Correlation is significant at the 0.01 level (2-tailed).

Note: the label "IRR" refers to entities identified as irrelevant.

Table C.16: Modified judgements, according to the given motivation.

Specified	Motivation	Modified
	Scenario 1 – iteration 1	
9115	These are close in time, but [9128] is better, since there are 4 close supermarkets.	9128
9123	[9123] leaves no time left to go shopping. (**Note**: 9123 specified as irrelevant)	9115
9115	same (**Note**: reference to the previous question whose this was the check. There the user specified 9128 as more relevant and motivated with "extra options are always a plus")	9128
None	I would never go to the supermarket after work. So [9115] would be my primary choice.	9115
	Scenario 2 – iteration 1	
9698	[9700] is closer & has more restaurants nearby	9700
675	Option [675] is not applicable and option [9698] fits all needs.	9698
Eq.	[9698] has hotels nearby in case of no vacancy; better walk a little more for a sure thing	9698
9828 as IRR	[9828] is the shorter walk, plus the distance between the pharmacies & restaurants of both hotels is the same, so why walk farther to the hotel when you can go to the closer one?	no IRR
	Scenario 2 – iteration 1	
746 as IRR, 711 as IRR	Option [711] is closed. (**Note**: 746 as more relevant)	711 as IRR
	Scenario 2 – iteration 1	
711 as more relevant, 746 as IRR	Option B was closed. Not an option. (**Note**: B is 746 but the motivation clearly refers to 711, which is the only one closed in this scenario)	746 as more relevant, 711 as IRR

Note: The letter A and B referring to the entities as referred to in the questionnaire have been replaced in the presented text with the identification number of the referred entities enclosed in round brackets.

Table C.17: Explanations given by the workers who specified entity 711 as more relevant in iteration 1 of Scenario 3.

Even though (711) is closed, there are more places to eat and pubs near by in just that location to choose from.
Although (711) is closed there are more options nearby to choose from.
Although (711)'s restaurant is closed, there are 18 other restaurants nearby so I'm sure we could find somewhere else to go, and there are a lot of pubs, too.
(711) fits better even though it is close. Having all the other restaurants and pubs close buy make it a better choice and it it closer to the bus station.
Even though the restaurant wasn't open it does fit the criteria
(711) might be closed, but it has a huge variety around it.
Even though (711) restaurant is closed there are many other possiblities in this area as opposed to (746) with only 2 restaurants.
Even though (711) is closed, the options in the area seem much more plentiful and better, with a higher chance of finding something that is open and that has space.
More options near by even though the initial restaurant is closed.
Although closed, there are many other options available nearby.
(711) might be closed right now, but it is closer to the other restaurants and pubs and also closer to the bus station. There is a pub right next doors by the looks of it, so it would be easy to just go there and have a few drinks while waiting to see if the restaurant will open.
Even though (711) is close it is so close to the others that it would give us a lot of alternatives.
Closed but meets all the other criteria
(711) has a greater variety close by, even if it is closed.
Option (711) has less walk time and more variety even though 1 minute longer walk time and a closed restaurant.
Even though (711) is closed, the options in the area seem much more plentiful and better, with a higher chance of finding something that is open and that has space.
I chose option (711) because even though it is closed it has more choices nearby than option (746).
(711) is still good because even if its closed but there are still 18 more other restaurant nearby.
(711) better fits my needs because, even though it's closed, it's near many more restaurants and pubs within walking distance.
I chose (711) because even though the restaurants closed the 18 will give good options.
I will still choose (711) because even if its closed but there are still more other restaurant to choose from so its not like "end of the world" to me.
(711) is closed but there are other restaurants nearby to choose from that would still be open, plus it's a shorter walk time.

Note: The letter A and B referring to the entities as referred to in the questionnaire have been replaced in the presented text with the identification number of the referred entities enclosed in round brackets.

Table C.18: Comparison between crowdsourced and computed ranks for Scenario 3. Three possible cases, considering entity 711 not irrelevant.

Correlation	*Baseline1*	*Baseline2*	*ScoreGR*	*GRBM25*
if 746 <711				
Kendall's tau	.057	.057	.686*	.400
p	>.05	>.05	<.05	>.05
if 746 = 711				
Kendall's tau	.087	.087	.667*	.377
p	>.05	>.05	<.05	>.05
if 746 >711 >747				
Kendall's tau	.114	.114	.629*	.343
p	>.05	>.05	<.05	>.05
if 746,747 >711 >724,7211				
Kendall's tau	.057	.057	.667*	.286
p	>.05	>.05	<.05	>.05

* Correlation is significant at the 0.05 level (2-tailed).

** Correlation is significant at the 0.01 level (2-tailed).

Bibliography

A. Abdalla and A. Frank. Combining Trip and Task Planning: How to Get from A to Passport. *Geographic Information Science*, pages 1–14, 2012.

A.N. Alazzawi, A.I. Abdelmoty, and C.B. Jones. An ontology of place and service types to facilitate place-affordance geographic information retrieval. In *Proceedings of the 6th Workshop on Geographic Information Retrieval*, page 4. ACM, 2010.

A.N. Alazzawi, A.I. Abdelmoty, and C.B. Jones. What can I do there? Towards the automatic discovery of place-related services and activities. *International Journal of Geographical Information Science*, 26(2):345–364, 2012.

J. Allan, B. Croft, A. Moffat, and M. Sanderson. Frontiers, challenges, and opportunities for information retrieval: Report from SWIRL 2012 the second strategic workshop on information retrieval in Lorne. In *ACM SIGIR Forum*, volume 46, pages 2–32. ACM, 2012.

O. Alonso and R. Baeza-Yates. Design and implementation of relevance assessments using crowdsourcing. *Advances in Information Retrieval*, pages 153–164, 2011.

O. Alonso and S. Mizzaro. Can we get rid of TREC assessors? Using Mechanical Turk for relevance assessment. In *Proceedings of the SIGIR 2009 Workshop on the Future of IR Evaluation*, pages 15–16, 2009.

O. Alonso, D.E. Rose, and B. Stewart. Crowdsourcing for relevance evaluation. In *ACM SIGIR Forum*, volume 42, pages 9–15. ACM, 2008.

A. Alves, F. Pereira, A. Biderman, and C. Ratti. Place enrichment by mining the web. *Ambient Intelligence*, pages 66–77, 2009.

A.O. Alves and F.C. Pereira. Making sense of location context. In *Proceedings of the 1st International Workshop on Context Discovery and Data Mining*, page 4. ACM, 2012.

L. Andrade and M.J. Silva. Relevance ranking for geographic IR. *ACM GIR*, 2006.

M.J. Barranco, J.M. Noguera, J. Castro, and L. Martínez. A Context-Aware Mobile Recommender System Based on Location and Trajectory. *Management Intelligent Systems*, pages 153–162, 2012.

C. L. Barry. User-defined relevance criteria: an exploratory study. *Journal of the American Society for Information Science*, 45(3):149–159, 1994. ISSN 1097-4571.

C. L. Barry and L. Schamber. Users' criteria for relevance evaluation: a cross-situational comparison. *Information Processing and Management*, 34(2-3):219–236, 1998.

P. Bereuter, R. Venkateswaran, and R. Weibel. The use of filters for adaptive mobile mapping scenarios. In *AGILE 2009 Workshop on Adaptation in Spatial Communication*, 2009.

T.J. Berners-Lee. The world-wide web. *Computer Networks and ISDN Systems*, 25(4-5): 454–459, 1992.

J. Bertin. *Semiology of graphics: diagrams, networks, maps.* University of Wisconsin press, 1983.

R. Bierig and A. Göker. Time location and interest: an empirical and user-centred study. In *Proceedings of the 1st international conference on Information interaction in context*, pages 79–87. ACM, October 2006 2006.

A. Bookstein. Relevance. *Journal of the American Society for Information Science*, 30 (5):269–73, 1979.

G. Bordogna and G. Psaila. Flexible Querying in Geo-Finder. In *Proceedings of the 3rd Italian Information Retrieval Workshop*, 2012.

P. Borlund. The concept of relevance in IR. *Journal of the American Society for information Science and Technology*, 54(10):913–925, 2003.

F.P. Boscoe, K.A. Henry, and M.S. Zdeb. A Nationwide Comparison of Driving Distance Versus Straight-Line Distance to Hospitals. *The Professional Geographer*, 64(2), 2012.

danah boyd and Kate Crawford. Six Provocations for Big Data. In *Decade in Internet Time: Symposium on the Dynamics of the Internet and Society*, 2011.

A.J. Brimicombe. Location-Based Services and Geographic Information Systems. *The Handbook of Geographic Information Science*, pages 581–595, 2008.

D. Buscaldi. Approaches to disambiguating toponyms. *SIGSPATIAL Special*, 3(2): 16–19, 2011.

G. Cai. GeoVSM: An integrated retrieval model for geographic information. In *Proceedings of the Second International Conference on Geographic Information Science*, pages 65–79, 2002.

G. Cai. Relevance ranking in Geographical Information Retrieval. *SIGSPATIAL Special*, 3(2):33–36, 2011.

X. Cao, G. Cong, and C.S. Jensen. Retrieving top-k prestige-based relevant spatial web objects. *Proceedings of the VLDB Endowment*, 3(1-2):373–384, 2010.

N.R. Chrisman. Part 2: Issues and Problems Relating to Cartographic Data Use, Exchange and Transfer: The Role Of Quality Information In The Long-Term Functioning Of A Geographic Information System. *Cartographica: The International Journal for Geographic Information and Geovisualization*, 21(2):79–88, 1984.

R.L. Cilibrasi and P.M.B. Vitanyi. The google similarity distance. *Knowledge and Data Engineering, IEEE Transactions on*, 19(3):370–383, 2007.

G. Cong, C.S. Jensen, and D. Wu. Efficient retrieval of the top-k most relevant spatial web objects. *Proceedings of the VLDB Endowment*, 2(1):337–348, 2009.

W.S. Cooper. A definition of relevance for information retrieval. *Information storage and retrieval*, 7(1):19–37, 1971.

P. Coppola, V. Della Mea, L. Di Gaspero, and S. Mizzaro. The concept of relevance in mobile and ubiquitous information access. In Fabio Crestani, Mark Dunlop, and Stefano Mizzaro, editors, *Mobile and Ubiquitous Information Access*, volume 2954 of *Lecture Notes in Computer Science*, pages 3–6. Springer Berlin / Heidelberg, 2004.

P. Coppola, V. Della Mea, L. Di Gaspero, D. Menegon, D. Mischis, S. Mizzaro, I. Scagnetto, and L. Vassena. The context-aware browser. *Intelligent Systems, IEEE*, 25(1):38–47, 2010.

E. Cosijn and P. Ingwersen. Dimensions of relevance. *Information Processing & Management*, 36(4):533–550, 2000.

H. Couclelis. Rethinking time geography in the information age. *Environment and planning. A*, 41(7):1556, 2009.

H. Couclelis, R.G. Golledge, N. Gale, and W. Tobler. Exploring the anchor-point hypothesis of spatial cognition. *Journal of Environmental Psychology*, 7(2):99–122, 1987. ISSN 0272-4944.

P. Crease. Time Geography in Support of Mobile Activity Planning. In *Proceedings of GISRUK Conference 2012*, 2012.

P. Crease and T. Reichenbacher. Designing Usable Cartographic Representations for Mobile Information Seeking. In *Proceedings of 25th International Cartographic Conference*, July 2011.

J. Crowley, J. Coutaz, G. Rey, and P. Reignier. Perceptual components for context aware computing. *UbiComp 2002: Ubiquitous Computing*, pages 117–134, 2002.

J.R. Curran, S. Clark, and J. Bos. Linguistically motivated large-scale NLP with C&C and Boxer. In *Proceedings of the 45th Annual Meeting of the ACL on Interactive*

Poster and Demonstration Sessions, pages 33–36. Association for Computational Linguistics, 2007.

C. da Costa Pereira, M. Dragoni, and G. Pasi. Multidimensional relevance: a new aggregation criterion. In Mohand Boughanem, Catherine Berrut, Josiane Mothe, and Chantal Soule-Dupuy, editors, *Advances in Information Retrieval*, volume 5478 of *Lecture Notes in Computer Science*, pages 264–275. Springer Berlin / Heidelberg, 2009.

C. da Costa Pereira, M. Dragoni, and G. Pasi. Multidimensional relevance: Prioritized aggregation in a personalized Information Retrieval setting. *Information processing & management*, 48(2):340–357, 2012.

S. De Bruin, A. Bregt, and M. Van De Ven. Assessing fitness for use: the expected value of spatial data sets. *International Journal of Geographical Information Science*, 15 (5):457–471, 2001.

S. De Sabbata. Criteria of geographic relevance. In *Proceedings of the 12th International Conference on Geographic Information Science*, 2010.

S. De Sabbata and T. Reichenbacher. A probabilistic model of geographic relevance. In *Proceedings of the 6th Workshop on Geographic Information Retrieval*, page 23. ACM, 2010.

S. De Sabbata and T. Reichenbacher. Criteria of geographic relevance: an experimental study. *International Journal of Geographical Information Science*, 26(8):1495–1520, 2012.

S. De Sabbata, O. Alonso, and S. Mizzaro. Classical vs. Crowdsourcing Surveys for Eliciting Geographic Relevance Criteria. In *Proceedings of the 3rd Italian Information Retrieval Wrokshop, Bari, Italy*, 2012.

A.K. Dey. Understanding and using context. *Personal and ubiquitous computing*, 5(1): 4–7, 2001.

S. Dodge, R. Weibel, and A.K. Lautenschütz. Towards a taxonomy of movement patterns. *Information Visualization*, 7(3-4):240–252, 2008.

S. Dodge, P. Laube, and R. Weibel. Movement similarity assessment using symbolic representation of trajectories. *International Journal of Geographical Information Science*, 26(9):1563–1588, 2012.

D. Dransch. Activity and Context-A Conceptual Framework for Mobile Geoservices. *Map-based mobile services, Springer Verlag, Berlin*, pages 31–44, 2005.

JJ Dujmovic. Extended continuous logic and the theory of complex criteria. *J. University of Belgrade, Series on Mathematics and Physics*, 537:197–216, 1975.

J.J. Dujmovic. Continuous preference logic for system evaluation. *Fuzzy Systems, IEEE Transactions on*, 15(6):1082–1099, 2007.

J.J. Dujmovic and H.L. Larsen. Generalized conjunction/disjunction. *International Journal of Approximate Reasoning*, 46(3):423–446, 2007.

A. D'Ulizia, F. Ferri, and P. Grifoni. A similarity assessment method for discovering and adapting business services. *International Journal of Computational Science and Engineering*, 5(2):97–109, 2010.

C. Eickhoff. Introduction to Crowdsourcing. *Delft University of Technology*, 2011.

C. Eickhoff and A. de Vries. How crowdsourcable is your task. In *Proceedings of the Workshop on Crowdsourcing for Search and Data Mining (CSDM) at the Fourth ACM International Conference on Web Search and Data Mining (WSDM)*, pages 11–14, 2011.

C. Emmanouilidis, R.A. Koutsiamanis, and A. Tasidou. Mobile Guides: Taxonomy of Architectures, Context Awareness, Technologies and Applications. *Journal of Network and Computer Applications*, 2012.

M. Ester, H.P. Kriegel, J. Sander, and X. Xu. A density-based algorithm for discovering clusters in large spatial databases with noise. In *Proceedings of the 2nd International Conference on Knowledge Discovery and Data mining*, volume 1996, pages 226–231. AAAI Press, 1996.

A. Fader, S. Soderland, and O. Etzioni. Identifying relations for open information extraction. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 1535–1545. Association for Computational Linguistics, 2011.

D. Fallows. *The Internet and daily life.* Pew Internet & American Life Project Washington, DC, 2004.

A.L. Felstiner. Working the crowd: employment and labor law in the crowdsourcing industry. 2010.

G. Fischer. User modeling in human–computer interaction. *User modeling and user-adapted interaction*, 11(1):65–86, 2001.

P. Frontiera, R. Larson, and J. Radke. A comparison of geometric approaches to assessing spatial similarity for GIR. *International Journal of Geographical Information Science*, 22(3):337–360, 2008.

F. Gey, R. Larson, M. Sanderson, K. Bischoff, T. Mandl, C. Womser-Hacker, D. Santos, P. Rocha, G. Di Nunzio, and N. Ferro. Geoclef 2006: The clef 2006 cross-language geographic information retrieval track overview. *Evaluation of Multilingual and Multi-modal Information Retrieval*, pages 852–876, 2007.

F. Giannotti, G. Giannotti, and D. Pedreschi. *Mobility, data mining, and privacy: geographic knowledge discovery.* Springer, 2008.

J.J. Gibson. *The ecological approach to visual perception.* Lawrence Erlbaum, 1986.

R.G. Golledge. The nature of geographic knowledge. *Annals of the Association of American Geographers*, 92(1):1–14, 2002. ISSN 1467-8306.

M. Goodchild. Volunteered geographic information. *GEOconnexion International Magazine*, pages 46–47, 2008.

M. Goodchild. Twenty years of progress: GIScience in 2010. *Journal of spatial information science*, (1):3–20, 2012.

M.F. Goodchild. Formalizing place in geographic information systems. *Communities, Neighborhoods, and Health*, pages 21–33, 2011.

S. Greenberg. Context as a dynamic construct. *Human-Computer Interaction*, 16(2-4): 257–268, 2001.

H. Greisdorf. Relevance: An interdisciplinary and information science perspective. *Informing Science*, 3(2):67–72, 2000.

D. Gunning, V.K. Chaudhri, and C. Welty. Introduction to the Special Issue on Question Answering. *AI Magazine*, 31(3):11–12, 2010.

S.C. Guptill, J.L. Morrison, and International Cartographic Association. *Elements of spatial data quality.* Elsevier Science Oxford, 1995.

T. Hägerstrand. What about people in Regional Science? *Papers in Regional Science*, 24(1):6–21, 1970. doi: 10.1007/BF01936872.

M. Haklay. How good is volunteered geographical information? A comparative study of OpenStreetMap and Ordnance Survey datasets. *Environment and planning. B, Planning & design*, 37(4):682, 2010.

D. Halpern and M.S. Nilan. A step toward shifting the research emphasis in information science from the system to the user: An empirical investigation of source-evaluation behavior information seeking and use. In *Proceedings of the 51st annual meeting of the American Society for Information Science*, volume 25, pages 169–176, 1988.

J. Han, M. Kamber, and A. K. H. Tung. *Spatial Clustering Methods in Data Mining: A Survey*, chapter 8, pages 188 – 217. Taylor and Francis, 2001.

C. Harris. Youre Hired! An Examination of Crowdsourcing Incentive Models in Human Resource Tasks. In *Proceedings of the Workshop on Crowdsourcing for Search and Data Mining (CSDM) at the Fourth ACM International Conference on Web Search and Data Mining (WSDM)*, pages 15–18, 2011.

T. Hashem and L. Kulik. "Don't trust anyone": Privacy protection for location-based services. *Pervasive and Mobile Computing*, 7(1):44–59, 2011.

E. Hauthal and D. Burghardt. Investigation and Development of Mobile Touristic Applications. *Advances in Location-Based Services*, pages 267–282, 2012.

S. Hirtle, S. Timpf, and T. Tenbrink. The effect of activity on relevance and granularity for navigation. *Spatial Information Theory*, pages 73–89, 2011.

J. Hong, E. Suh, and S.J. Kim. Context-aware systems: A literature review and classification. *Expert Systems with Applications*, 36(4):8509–8522, 2009.

H. Huang and G. Gartner. Using activity theory to identify relevant context parameters. *Location Based Services and TeleCartography II*, pages 35–45, 2009.

H. Huang and G. Gartner. Using Context-Aware Collaborative Filtering for POI Recommendations in Mobile Guides. *Advances in Location-Based Services*, pages 131–147, 2012.

Y. Huang, S. Shekhar, and H. Xiong. Discovering colocation patterns from spatial data sets: a general approach. *Knowledge and Data Engineering, IEEE Transactions on*, 16(12):1472–1485, 2004.

P. Ipeirotis. Mechanical turk: The demographics. *A Computer Scientist in a Business School*, 2008.

P. Ipeirotis. Demographics of mechanical turk. *Center for Digital Economy Research, NYU Stern School of Business, Working paper*, 2010.

K. Janowicz, S. Scheider, T. Pehle, and G. Hart. Geospatial semantics and linked spatiotemporal data–Past, present, and future. *Semantic Web*, 3(4):321–332, 2012.

B. Jiang and X. Yao. Location-based services and GIS in perspective. *Computers, Environment and Urban Systems*, 30(6):712–725, 2006.

Y. Jiang, L. Wang, Y. Lu, and H. Chen. Discovering both positive and negative colocation rules from spatial data sets. In *Software Engineering and Data Mining (SEDM), 2010 2nd International Conference on*, pages 398–403. IEEE, 2010.

T. Joachims. Optimizing search engines using clickthrough data. In *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 133–142. ACM, 2002.

C.B. Jones and R.S. Purves. Geographical information retrieval. *International Journal of Geographical Information Science*, 22(3):219–228, 2008a.

C.B. Jones and R.S. Purves. Web-Based Geographic Information Systems. *The handbook of geographic information science*, pages 559–580, 2008b.

C.B. Jones, R.S. Purves, P.D. Clough, and H. Joho. Modelling vague places with knowledge from the Web. *International Journal of Geographical Information Science*, 22 (10):1045–1065, 2008.

G.J.F. Jones and P.J. Brown. Context-aware retrieval for ubiquitous computing environments. *Mobile and ubiquitous information access*, pages 371–374, 2004.

T. Jordan, M. Raubal, B. Gartrell, and M. Egenhofer. An affordance-based model of place in GIS. In *Proceedings of 8th International Symposium on Spatial Data Handling*, pages 98–109. Citeseer, July 1998 1998.

M. Kaenampornpan and E. O'Neill. Modelling context: An activity theory approach. *Ambient Intelligence*, pages 367–374, 2004.

K. Kageura, T. Koyama, M. Yoshioka, A. Takasu, T. Nozue, and K. Tsuji. NACSIS corpus project for IR and terminological research. In *Natural Language Processing Pacific Rim Symposium*, volume 97, pages 2–5. Citeseer, 1997.

N. Kando, K. Kuriyama, and T. Nozue. NACSIS test collection workshop. In *Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval*, pages 299–300. ACM, 1999.

G.M. Kapitsaki, G.N. Prezerakos, N.D. Tselikas, and I.S. Venieris. Context-aware service engineering: A survey. *Journal of Systems and Software*, 82(8):1285–1297, 2009.

V. Kaptelinin and B.A. Nardi. Activity theory: basic concepts and applications. In *CHI'97 extended abstracts on Human factors in computing systems: looking to the future*, pages 158–159. ACM, 1997.

V. Kaptelinin and BA Nardi. Activity theory in a nutshell. *Acting with Technology: Activity Theory and Interaction Design*, pages 29–72, 2006.

M.G. Kendall. A new measure of rank correlation. *Biometrika*, 30(1/2):81–93, 1938.

A. Kofod-Petersen and J. Cassens. Using activity theory to model context awareness. *Modeling and Retrieval of Context*, pages 1–17, 2006.

K. Koperski and J. Han. Discovery of spatial association rules in geographic information databases. In Max Egenhofer and John Herring, editors, *Advances in Spatial Databases*, volume 951 of *Lecture Notes in Computer Science*, pages 47–66. Springer Berlin / Heidelberg, 1995.

B. Kuijpers, H.J. Miller, and W. Othman. Kinetic space-time prisms. In *Proceedings of the 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, pages 162–170. ACM, 2011.

R.R. Larson. Geographic information retrieval and spatial browsing. In *GIS and Libraries Patrons, Maps and Spatial Information*. Graduate School of Library and Information Science, University of Illinois at Urbana-Champaign, 1996.

R.R. Larson. Ranking approaches for GIR. *SIGSPATIAL Special*, 3(2):37–41, 2011.

R.R. Larson and P. Frontiera. Spatial ranking methods for geographic information retrieval (GIR) in digital libraries. *Research and Advanced Technology for Digital Libraries*, pages 45–56, 2004.

M. Laukkanen, K. Viljanen, M. Apiola, P. Lindgren, and E. Hyvönen. Towards ontology-based yellow page services. In *Proceedings of WWW2004 Workshop on Application Design, Development, and Implementation Issues in the Semantic Web, New York, USA*, 2004.

J.L. Leidner and M.D. Lieberman. Detecting geographical references in the form of place names and associated spatial natural language. *SIGSPATIAL Special*, 3(2): 5–11, 2011.

J. Leveling. Challenges for Indexing in GIR. *SIGSPATIAL Special*, 3(2):29–32, 2011.

H. Lieberman, A. Faaborg, J. Espinosa, and T. Stocky. Commonsense on the go. *BT Technology Journal*, 22(4):241–252, 2004.

D. Lin. An information-theoretic definition of similarity. In *Proceedings of the 15th international conference on Machine Learning*, volume 1, pages 296–304. San Francisco, 1998.

H. Liu and P. Singh. ConceptNet—a practical commonsense reasoning tool-kit. *BT technology journal*, 22(4):211–226, 2004.

B. Luyt, A.C. Tay, H.T. Lim, and K.H. Cheng. Novelty and topicality in interactive information retrieval. *J. Am. Soc. Inf. Sci. Technol.*, 59(2):201–215, 2008.

K. Lynch. The city image and its elements. *The image of the city*, pages 46–90, 1960.

T. Mandl. Evaluating GIR: geography-oriented or user-oriented? *SIGSPATIAL Special*, 3(2):42–45, 2011.

T. Mandl, F. Gey, G. Di Nunzio, N. Ferro, R. Larson, M. Sanderson, D. Santos, C. Womser-Hacker, and X. Xie. Geoclef 2007: the clef 2007 cross-language geographic information retrieval track overview. *Advances in Multilingual and Multimodal Information Retrieval*, pages 745–772, 2008.

T. Mandl, P. Carvalho, G.M. Di Nunzio, F. Gey, R.R. Larson, D. Santos, and C. Womser-Hacker. *GeoCLEF 2008: The CLEF 2008 Cross-Language Geographic Information Retrieval Track Overview*, volume 5706 of *Lecture Notes in Computer Science*, pages 808–821. Springer, 2009. doi: 10.1007/978-3-642-04447-2_106.

C.D. Manning, P. Raghavan, and H. Schutze. *Introduction to information retrieval*, volume 1. Cambridge University Press Cambridge, 2008.

D. Manov, A. Kiryakov, B. Popov, K. Bontcheva, D. Maynard, and H. Cunningham. Experiments with geographic knowledge for information extraction. In *Proceedings of the HLT-NAACL 2003 workshop on Analysis of geographic references-Volume 1*, pages 1–9. Association for Computational Linguistics, 2003.

M. Marchiori. The quest for correct information on the web: Hyper search engines. *Computer Networks and ISDN Systems*, 29(8-13):1225–1235, 1997.

D.M. Mark. Geographic information science: Defining the field. *Foundations of geographic information science*, pages 3–18, 2003.

M.E. Maron and J.L. Kuhns. On relevance, probabilistic indexing and information retrieval. *Journal of the ACM (JACM)*, 7(3):216–244, 1960.

P. Marsden. Crowdsourcing. *Contagious Magazine*, 18:24–28, 2009.

B. Martins, M.J. Silva, and L. Andrade. Indexing and ranking in Geo-IR systems. In *Proceedings of the 2005 workshop on Geographic information retrieval*, pages 31–34. ACM, 2005.

W. Mason and S. Suri. Conducting behavioral research on Amazonś Mechanical Turk. *Behavior research methods*, pages 1–23, 2011.

R.M.C. McCreadie, C. Macdonald, and I. Ounis. Crowdsourcing a news query classification dataset. In *Proceedings of the ACM SIGIR 2010 workshop on crowdsourcing for search evaluation (CSE 2010)*, pages 31–38, 2010.

L. Meng. Ego centres of mobile users and egocentric map design. *Mapbased Mobile Services*, pages 87–105, 2005.

G.A. Miller et al. WordNet: a lexical database for English. *Communications of the ACM*, 38(11):39–41, 1995.

H.J. Miller. A measurement theory for time geography. *Geographical analysis*, 37(1): 17–45, 2005a.

H.J. Miller. Necessary space- time conditions for human interaction. *Environment and Planning B: Planning and Design*, 32(3):381–401, 2005b.

H.J. Miller and S.A. Bridwell. A field-based theory for time geography. *Annals of the Association of American Geographers*, 99(1):49–75, 2009. ISSN 0004-5608.

H.J. Miller and J. Han. *Geographic data mining and knowledge discovery*. CRC, first edition edition, 2001.

H.J. Miller and J. Han. *Geographic data mining and knowledge discovery*. CRC, second edition edition, 2009.

S. Mizzaro. Relevance: The whole history. *Journal of the American society for information science*, 48(9):810–832, 1997a.

S. Mizzaro. Relevance: the whole history. *Journal of the American Society for Information Science*, 48(9):810–832, 1997b.

S. Mizzaro. How many relevances in information retrieval? *Interacting with computers*, 10(3):303–320, 1998.

S. Mizzaro and L. Vassena. A social approach to context-aware retrieval. *World Wide Web*, pages 1–29, 2011.

P. Mooney and P. Corcoran. Using OSM for LBS–An Analysis of Changes to Attributes of Spatial Objects. *Advances in Location-Based Services*, pages 165–179, 2012.

J.L. Morrison. Spatial data quality. *Elements of spatial data quality*, pages 1–12, 1995.

D. Mountain. *Exploring mobile trajectories: An investigation of individual spatial behaviour and geographic filters for information retrieval*. PhD thesis, City University, London, 2005.

D. Mountain and A. Macfarlane. Geographic information retrieval in a mobile environment: evaluating the needs of mobile individuals. *Journal of Information Science*, 33(5):515–530, 2007.

B.A. Nardi. Studying context: A comparison of activity theory, situated action models, and distributed cognition. *Context and consciousness: Activity theory and human-computer interaction*, pages 69–102, 1996.

M.S. Nilan, R.P. Peek, and H.W. Snyder. A methodology for tapping user evaluation behaviors: An exploration of users strategy, source and information evaluating. In *Proceedings of the 51st ASIS Annual Meeting*, volume 25, pages 152–159, 1988.

V. C. Ostuni, T. Di Noia, R. Mirizzi, D. Romito, and E. Di Sciascio. Cinemappy: a Context-aware Mobile App for Movie Recommendations boosted by DBpedia. In *1st International Workshop on Semantic Technologies meet Recommender Systems & Big Data (SeRSy 2012)*. CEUR-WS, 2012. URL `http://sisinflab.poliba.it/sisinflab/publications/2012/ODMRD12`.

D. O'Sullivan and D.J. Unwin. *Geographic information analysis*. John Wiley & Sons Inc, 2003.

S. Overell. The problem of place name ambiguity. *SIGSPATIAL Special*, 3(2):12–15, 2011.

L. Page, S. Brin, R. Motwani, and T. Winograd. The PageRank citation ranking: Bringing order to the web. 1999.

D. Palacio, C. Sallaberry, M. Gaio, et al. Normalizing spatial information to better combine criteria in geographical information retrieval. In *ECIR GIIW: 31st European Conference on Information Retrieval, Geographic Information on the Internet Workshop (GIIW)*, pages 37–49, 2009.

D. Palacio, G. Cabanac, C. Sallaberry, and G. Hubert. On the evaluation of Geographic Information Retrieval systems. *International Journal on Digital Libraries*, 11(2): 91–109, 2010.

D. Palacio, G. Cabanac, C. Sallaberry, and G. Hubert. On the evaluation of Geographic Information Retrieval systems. *International Journal on Digital Libraries*, pages 1–19, 2011.

F. Pereira, A. Alves, J. Oliveirinha, and A. Biderman. Perspectives on semantics of the place from online resources. In *Semantic Computing, 2009. ICSC'09. IEEE International Conference on*, pages 215–220. IEEE, 2009.

C. Peters and M. Braschler. European research letter: Cross-language system evaluation: The CLEF campaigns. *Journal of the American Society for Information Science and Technology*, 52(12):1067–1072, 2001.

D.J. Peuquet. A conceptual framework and comparison of spatial data models. *Cartographica: The International Journal for Geographic Information and Geovisualization*, 21(4):66–113, 1984.

D.J. Peuquet. *Representations of space and time*. Guilford Press, 2002.

S.M. Pollock. Measures for the comparison of information retrieval systems. *American Documentation*, 19(4):387–397, 1968.

P. Pombinho, M.B. Carmo, and A.P. Afonso. Context aware point of interest adaptive recommendation. In *Proceedings of the 2nd Workshop on Context-awareness in Retrieval and Recommendation*, pages 30–33. ACM, 2012.

A. Pred. The choreography of existence: comments on Hägerstrand's time-geography and its usefulness. *Economic geography*, pages 207–221, 1977.

R. Purves and C. Jones. Geographic Information Retrieval. *SIGSPATIAL Special*, 3(2): 2–4, 2011.

R. Purves, A.K. Syed, B. Yang, and R. Weibel. A cartographic visualisation interface for spatial information retrieval. In *International Cartographic Conference (ICC), La Coruna, Spain*, 2005.

R. Purves, P. Clough, C.B. Jones, A. Arampatzis, B. Bucher, D. Finch, G. Fu, H. Joho, A.K. Syed, S. Vaid, et al. The design and implementation of SPIRIT: a spatially aware search engine for information retrieval on the Internet. *International Journal of Geographical Information Science*, 21(7):717–745, 2007.

S. Putz. Interactive information services using World-Wide Web hypertext. *Computer Networks and ISDN Systems*, 27(2):273–280, 1994.

J. Raper. Geographic relevance. *Journal of Documentation*, 63(6):836–852, 2007.

J. Raper, G. Gartner, H. Karimi, and C. Rizos. Applications of location–based services: a selected review. *Journal of Location Based Services*, 1(2):89–111, 2007a.

J. Raper, G. Gartner, H. Karimi, and C. Rizos. A critical evaluation of location based services and their potential. *Journal of Location Based Services*, 1(1):5–45, 2007b.

A.M. Rashid, I. Albert, D. Cosley, S.K. Lam, S.M. McNee, J.A. Konstan, and J. Riedl. Getting to know you: learning new user preferences in recommender systems. In *Proceedings of the 7th international conference on Intelligent user interfaces*, pages 127–134. ACM, 2002.

M. Raubal and I. Panov. A formal model for mobile map adaptation. *Location Based Services and TeleCartography II*, pages 11–34, 2009.

M. Raubal, H.J. Miller, and S. Bridwell. User-centred time geography for location-based services. *Geografiska Annaler Series B Human Geography*, 86(4):245–265, 2004.

T. Reichenbacher. The world in your pocket-towards a mobile cartography. In *Proceedings of the 20th International Cartographic Conference, Beijing, China*, volume 4, pages 2514–2521. Citeseer, 2001.

T. Reichenbacher. *Mobile cartography: adaptive visualisation of geographic information on mobile devices*. Verlag Dr. Hut, 2004.

T. Reichenbacher. The Importance of being relevant. In *Proceedings XXII International Cartographic Conference*, July 2005 2005a.

T. Reichenbacher. Adaptive egocentric maps for mobile users. *Map-based mobile services: theories, methods and implementations*, 1:141, 2005b.

T. Reichenbacher. The concept of relevance in mobile maps. In Georg Gartner, William Cartwright, Michael P. Peterson, William Cartwright, Georg Gartner, Liqiu Meng, and Michael P. Peterson, editors, *Location Based Services and TeleCartography*, Lecture Notes in Geoinformation and Cartography, pages 231–246. Springer Berlin / Heidelberg, 2007.

T. Reichenbacher. Mobile Usage and Adaptive Visualization. In Encyclopedia of GIS, Part 16, pp. 677-682. Shekhar S. and Xiong H. (Eds), Heidelberg, Springer Verlag, 2008.

T. Reichenbacher. Geographic relevance in mobile services. In *Proceedings of the 2nd International Workshop on Location and the Web*. ACM, April 2009 2009.

T. Reichenbacher and S. De Sabbata. Geographic relevance: different notions of geographies and relevancies. *SIGSPATIAL Special*, 3(2):67–70, 2011.

T. Reichenbacher, P. Crease, and S. De Sabbata. The concept of geographic relevance. In *Proceedings of the 6th International Symposium on LBS & TeleCartography* , 2009.

J. Riegelsberger, M. Lee, and S. Lederer. A room with a view: understanding users' stages in picking a hotel online. In *Proceedings of the 2012 ACM annual conference extended abstracts on Human Factors in Computing Systems Extended Abstracts*, pages 713–716. ACM, 2012.

S.E. Robertson and K.S. Jones. Relevance weighting of search terms. *Journal of the American Society for Information science*, 27(3):129–146, 1976.

J. Ross, L. Irani, M.S. Silberman, A. Zaldivar, and B. Tomlinson. Who are the crowdworkers?: shifting demographics in Mechanical Turk. In *Proceedings of the 28th of the international conference on Human factors in computing systems*, pages 2863–2872. ACM, April 2010 2010.

S.J. Russell, P. Norvig, J.F. Canny, J.M. Malik, and D.D. Edwards. *Artificial intelligence: a modern approach*, volume 2. Prentice hall Englewood Cliffs, NJ, 1995.

N. Saiph Savage, M. Baranski, N. Elva Chavez, and T. Höllerer. I'm feeling loco: A location based context aware recommendation system. *Advances in Location-Based Services*, pages 37–54, 2012.

G. Salton. The SMART retrieval system: experiments in automatic document processing. 1971.

J. Sander, M. Ester, H.P. Kriegel, and X. Xu. Density-based clustering in spatial databases: The algorithm gdbscan and its applications. *Data Mining and Knowledge Discovery*, 2(2):169–194, 1998.

M. Sanderson. Test Collection Based Evaluation of Information Retrieval Systems. *Information Retrieval*, 4(4):247–375, 2010.

D. Santos and L. Cabral. GikiCLEF: Expectations and lessons learned. *Multilingual Information Access Evaluation I. Text Retrieval Experiments*, pages 212–222, 2010.

D. Santos and L.M. Cabral. GikiCLEF: Crosscultural issues in an international setting: asking non-English-centered questions to Wikipedia. In *Cross Language Evaluation Forum: Working notes for CLEF*, volume 30, 2009.

T. Saracevic. Relevance reconsidered. In *Proceedings of the 2nd Conference on Conceptions of Library and Information Science*, pages 201–218. Royal School of Librarianship, October 1996 1996.

T. Saracevic. Relevance: a review of the literature and a framework for thinking on the notion in information science. Part II: nature and manifestations of relevance. *Journal of the American Society for Information Science and Technology*, 58(13): 1915–1933, 2007.

T. Saracevic and P. Kantor. A study of information seeking and retrieving. II. Users, questions, and effectiveness. *Journal of the American Society for Information Science*, 39(3):177–196, 1988a.

T. Saracevic and P. Kantor. A study of information seeking and retrieving. III. Searchers, searches, and overlap. *Journal of the American Society for Information Science*, 39 (3):197–216, 1988b.

T. Saracevic, P. Kantor, A.Y. Chamis, and D. Trivison. A study of information seeking and retrieving: I. Background and methodology. *Journal of the American Society for Information Science*, 39(3):161–176, 1988.

R. Savolainen and J. Kari. User-defined relevance criteria in web searching. *Journal of Documentation*, 62(6):685–707, 2006.

L. Schamber. Users' criteria for evaluation in multimedia information seeking and use situations. *Information Processing and Management*, 30(2), 1991.

L. Schamber, M.B. Eisenberg, and M.S. Nilan. A re-examination of relevance: toward a dynamic, situational definition. *Information processing & management*, 26(6): 755–776, 1990.

F. Schilder, Y. Versley, and C. Habel. Extracting spatial information: grounding, classifying and linking spatial expressions. In *Proceedings of the Workshop on Geographic Information Retrieval at SIGIR 2004*, 2004.

B. Schilit, N. Adams, and R. Want. Context-aware computing applications. In *Mobile Computing Systems and Applications, 1994. WMCSA 1994. First Workshop on*, pages 85–90. IEEE, 1994.

C. Schlieder. Modeling collaborative semantics with a geographic recommender. *Advances in Conceptual Modeling–Foundations and Applications*, pages 338–347, 2007.

C. Schlieder and A. Henrich. Spatial grounding with vague place models. *SIGSPATIAL Special*, 3(2):20–23, 2011.

C. Schlieder and D. Kremer. Visiting the same place but seeing different things: Place models of touristic behavior. *Visibility in Information Spaces and in Geographic Environments*, page 15, 2011.

C. Schlieder and C. Matyas. Photographing a city: An analysis of place concepts based on spatial choices. *Spatial Cognition & Computation*, 9(3):212–228, 2009.

A. Schmidt, M. Beigl, and H.W. Gellersen. There is more to context than location. *Computers & Graphics*, 23(6):893–901, 1999.

S. Schockaert. Vague regions in geographic information retrieval. *SIGSPATIAL Special*, 3(2):24–28, 2011.

S. Schockaert and M. De Cock. Neighborhood restrictions in geographic IR. In *Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 167–174. ACM, 2007.

I. Seifert. Collaborative assistance with spatio-temporal planning problems. *Spatial Cognition V Reasoning, Action, Interaction*, pages 90–106, 2007.

X. Shen, B. Tan, and C.X. Zhai. Context-sensitive information retrieval using implicit feedback. In *Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 43–50. ACM, 2005.

N. Shiode, C. Li, M. Batty, PA Longley, and D. Maguire. The impact and penetration of location-based services, 2002.

P. Singh, T. Lin, E. Mueller, G. Lim, T. Perkins, and W. Li Zhu. Open Mind Common Sense: Knowledge acquisition from the general public. *On the Move to Meaningful Internet Systems 2002: CoopIS, DOA, and ODBASE*, pages 1223–1237, 2002.

C. Skouteli, G. Samaras, and E. Pitoura. Concept-based discovery of mobile services. In *Proceedings of the 6th international conference on Mobile data management*, pages 257–261. ACM, 2005.

M. Soleymani and M. Larson. Crowdsourcing for affective annotation of video: Development of a viewer-reported boredom corpus. In *Proceedings of the ACM SIGIR 2010 workshop on crowdsourcing for search evaluation (CSE 2010)*, pages 4–8, 2010.

K. Spärck Jones. A statistical interpretation of term specificity and its application in retrieval. *Journal of documentation*, 28(1):11–21, 1972.

K. Spärck Jones, S. Walker, and S.E. Robertson. A probabilistic model of information retrieval: development and comparative experiments:: Part 2. *Information Processing & Management*, 36(6):809–840, 2000.

R. Straumann and R. Purves. Delineation of valleys and valley floors. *Geographic Information Science*, pages 320–336, 2008.

O. Swienty and T. Reichenbacher. Relevanz und Kognition in der mobilen Kartographie. *Aktuelle Entwicklungen in Geoinformation und Visualisierung. GEOVIS*, pages 5–6, 2006.

O. Swienty, T. Reichenbacher, S. Reppermund, and J. Zihl. The role of relevance and cognition in attention-guiding geovisualisation. *Cartographic Journal, The*, 45(3): 227–238, 2008.

W.R. Tobler. A computer movie simulating urban growth in the Detroit region. *Economic geography*, 46:234–240, 1970.

A.L. Towers and B.M. Gittings. Earthquake monitoring and prediction: a case study of GIS data integration using the Internet. *Innovations in GIS*, 2:233–243, 1995.

J. Urbano, J. Morato, M. Marrero, and D. Martín. Crowdsourcing preference judgments for evaluation of music similarity tasks. In *ACM SIGIR Workshop on Crowdsourcing for Search Evaluation*, pages 9–16, 2010.

S. Vaid, C. Jones, H. Joho, and M. Sanderson. Spatio-textual indexing for geographical search on the web. *Advances in Spatial and Temporal Databases*, pages 923–923, 2005.

M. Van Kreveld, I. Reinbacher, A. Arampatzis, and R. Van Zwol. Multi-Dimensional Scattered Ranking Methods for Geographic Information Retrieval. *Geoinformatica*, 9(1):61–84, 2005.

C.J. van Rijsbergen. *Information retrieval.* Butterworths, London, 2nd edition edition, 1979.

P. Venetis, H. Gonzalez, C.S. Jensen, and A. Halevy. Hyper-local, directions-based ranking of places. *Proc. VLDB Endow.*, 4(5):290–301, February 2011. ISSN 2150-8097. URL `http://dl.acm.org/citation.cfm?id=1952376.1952379`.

E. Voorhees, D.K. Harman, National Institute of Standards, and Technology (US). *TREC: Experiment and evaluation in information retrieval*, volume 63. MIT press Cambridge, 2005.

E.M. Voorhees. Query expansion using lexical-semantic relations. In *Proceedings of the 17th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 61–69. Springer-Verlag New York, Inc., 1994.

P. Wilson. Situational relevance. *Information storage and retrieval*, 9(8):457–471, 1973.

D. Wu, G. Cong, and C.S. Jensen. A framework for efficient spatial web object retrieval. *The VLDB Journal*, pages 1–26, 2012.

J. Xu and W.B. Croft. Query expansion using local and global document analysis. In *Proceedings of the 19th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 4–11. ACM, 1996.

Y. C. Xu and Z. Chen. Relevance judgment: What do information users consider beyond topicality? *Journal of the American Society for Information Science and Technology*, 57(7):961–973, 2006.

S. Yu, M.A. Aufaure, N. Cullot, and S. Spaccapietra. Location-based spatial modelling using ontology. In *AGILE 2003: 6th AGILE Conference on Geographic Information Science*, page 293. PPUR, 2003.

M.C. Yuen, I. King, and K.S. Leung. A survey of crowdsourcing systems. In *Privacy, Security, Risk and Trust (PASSAT), 2011 IEEE Third International Conference on and 2011 IEEE Third International Confernece on Social Computing (SocialCom)*, pages 766–773. IEEE, 2011.

Z. Zhang, A.L. Gentile, and F. Ciravegna. Recent advances in methods of lexical semantic relatedness–a survey. *Natural Language Engineering*, 2012.

A. Zipf. Die Relevanz von Geoobjekten in Fokuskarten. In *Proceedings of Symposium für Angewandte Geographische Informationstechnologie*, pages 567–576. Heidelberg: Wichmann, July 2003 2003.

A. Zipf and M. Jöst. Implementing adaptive mobile GI services based on ontologies:: Examples from pedestrian navigation support. *Computers, environment and urban systems*, 30(6):784–798, 2006.

A. Zipf and K.F. Richter. Using focus maps to ease map reading. *Künstliche Intelligenz*, 4(02):35–37, 2002.